



# Approches Métagénomiques en Microbiologie du Sol: Optimisation des Techniques Conventionnelles et Développement d'une Approche Alternative par Capture de Gènes dans une Bactérie Réceptrice.

Nathalie Lombard

## ► To cite this version:

Nathalie Lombard. Approches Métagénomiques en Microbiologie du Sol: Optimisation des Techniques Conventionnelles et Développement d'une Approche Alternative par Capture de Gènes dans une Bactérie Réceptrice.. Sciences de l'environnement. UCBL, 2007. Français. NNT: 120-2007 . tel-01211276

**HAL Id: tel-01211276**

**<https://theses.hal.science/tel-01211276>**

Submitted on 4 Oct 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**THESE**

Présentée devant

**L'UNIVERSITE CLAUDE BERNARD - LYON 1**

pour l'obtention du

**DIPLOME DE DOCTORAT**

(Arrêté du 7 août 2006)

Présentée et soutenue à huis-clos le 13 Juillet 2007

par

**Nathalie LOMBARD**

---

**APPROCHES METAGENOMIQUES EN MICROBIOLOGIE DU SOL :  
OPTIMISATION DES TECHNIQUES CONVENTIONNELLES  
ET DEVELOPPEMENT D'UNE APPROCHE ALTERNATIVE  
PAR CAPTURE DE GENES DANS UNE BACTERIE RECEPTRICE.**

---

Directeur de thèse : Pascal SIMONET

**JURY :**

Dietrich J.	Chercheur IFREMER, Université Montpellier 2, Montpellier	Rapporteur
Frostegard A.	Professeur, Université Agricole de Norvège, Aas	Rapporteur
Peyret P.	Professeur, Université Blaise Pascal, Clermont-Ferrand	Rapporteur
Bontems F.	Chargé de Recherche CNRS, Ecole Polytechnique, Palaiseau	Examineur
Lazzaroni J.C.	Directeur de Recherche CNRS, Université Lyon 1, Villeurbanne	Examineur
Nalin R.	Directeur Général, LibraGen S.A., Toulouse	Examineur
Rohr R.	Professeur, Université Lyon 1, Villeurbanne	Examineur
Simonet P.	Directeur de Recherche CNRS, Ecole Centrale, Ecully	Directeur

# UNIVERSITE CLAUDE BERNARD - LYON I

## **Président de l'Université**

Vice-Président du Conseil Scientifique

Vice-Président du Conseil d'Administration

Vice-Président du Conseil des Etudes et de la Vie Universitaire

**Secrétaire Général**

**M. le Professeur L. COLLET**

M. le Professeur J.F. MORNEX

M. le Professeur J. LIETO

M. le Professeur D. SIMON

**M. G. GAY**

## ***SECTEUR SANTE***

### ***Composantes***

UFR de Médecine Lyon R.T.H. Laënnec

UFR de Médecine Lyon Grange-Blanche

UFR de Médecine Lyon-Nord

UFR de Médecine Lyon-Sud

UFR d'Odontologie

Institut des Sciences Pharmaceutiques et Biologiques

Institut Techniques de Réadaptation

Département de Formation et Centre de Recherche en Biologie Humaine

Directeur : M. le Professeur D. VITAL-DURAND

Directeur : M. le Professeur X. MARTIN

Directeur : M. le Professeur F. MAUGUIERE

Directeur : M. le Professeur F.N. GILLY

Directeur : M. O. ROBIN

Directeur : M. le Professeur F. LOCHER

Directeur : M. le Professeur MATILLON

Directeur : M. le Professeur P. FARGE

## ***SECTEUR SCIENCES***

### ***Composantes***

UFR de Physique

UFR de Biologie

UFR de Mécanique

UFR de Génie Electrique et des Procédés

UFR Sciences de la Terre

UFR de Mathématiques

UFR d'Informatique

UFR de Chimie Biochimie

UFR STAPS

Observatoire de Lyon

Institut des Sciences et des Techniques de l'Ingénieur de Lyon

IUT A

IUT B

Institut de Science Financière et d'Assurances

Directeur : M. le Professeur A. HOAREAU

Directeur : M. le Professeur H. PINON

Directeur : M. le Professeur H. BEN HADID

Directeur : M. le Professeur A. BRIGUET

Directeur : M. le Professeur P. HANTZPERGUE

Directeur : M. le Professeur M. CHAMARIE

Directeur : M. le Professeur M. EGEA

Directeur : Mme. le Professeur H. PARROT

Directeur : M. le Professeur R. MASSARELLI

Directeur : M. le Professeur R. BACON

Directeur : M. le Professeur J. LIETO

Directeur : M. le Professeur M. C. COULET

Directeur : M. le Professeur R. LAMARTINE

Directeur : M. le Professeur J.C. AUGROS

# Remerciements

*Ce travail de thèse a été rendu possible grâce au financement accordé par la Région Rhône-Alpes dans le cadre « thématiques prioritaires ».*

*A ce titre, je souhaiterais vivement remercier Pascal Simonet, pour m'avoir proposé de travailler sur ce projet... Malgré les risques encourus, ce travail fut passionnant. Bien que les échecs rencontrés furent parfois décevants, ils n'en furent pas moins très stimulants, sources de réflexion et de challenges... Les réussites n'en furent que plus savourées ! Je tiens d'ailleurs à le remercier pour la confiance qu'il m'a accordée dans de nombreuses situations, son soutien immodéré, ainsi que pour son encadrement, les nombreuses participations à des congrès et les possibilités de rencontres offertes. Ce fut une véritable chance pour moi et je lui en suis très reconnaissante ...*

*Je profite de ces quelques lignes pour remercier toutes les personnes qui m'ont fortement aidées dans ce travail de thèse, car ce travail je n'aurais pu le réaliser seule... Aussi je tenais à mettre en valeur la contribution de chacun dans ce travail :*

*Je tiens tout d'abord à remercier Jean Claude Lazzaroni, pour son énorme contribution dans l'avancement du projet. Ses conseils furent un vrai tremplin dans la réalisation de ce travail. Il est notamment intervenu dans toutes les modifications chromosomiques de E. coli. Merci pour tout le temps accordé aussi bien dans les discussions, que dans les recherches, et expérimentations. Son aide fut et reste très précieuse.*

*Je voudrais également remercier Xiaojun Zhang, dont la collaboration dans ce projet fut très fructueuse. Il est plus particulièrement intervenu dans les premières expérimentations in vitro du système de contre sélection développé au cours de cette thèse. Sa contribution dans ce travail a rendu possible la présentation des premiers résultats obtenus. Xiaojun, your arrival in France was a new breath in experiences. I really enjoy working with you! You just stayed some months, but it was really enriching, and productive! Thank you for all...*

*Merci également à Aurélie Faugier, Jordan Madic et Lorris Niard, qui ont participé dans différentes étapes de construction génétiques contenant des gènes toxiques. Jordan a notamment eu la chance ( ? ) d'intervenir lors de nos premiers déboires avec les gènes toxiques, les optimisations techniques et les premières avancées expérimentales, Lorris ensuite, puis Aurélie qui fut d'un grand secours pour la finalisation du système contenant deux gènes toxiques.*

*Un merci très chaleureux à Hélène Bertrand qui m'a initiée à la métagénomique et avec qui j'ai eu la chance de travailler sur l'extraction des cellules de l'environnement sol. Un merci tout particulier à Jean Michel Monier, pour qui la microscopie (et l'anglais) n'a aucun secret et avec qui j'ai eu le plaisir de travailler sur la détection et l'isolement d'espèces bactériennes (électro) transformables ! Merci également à Patrick Mavingui, Pierre Pujic, et Franck Bertolla pour toutes les discussions fructueuses, tant sur les mystères de la biologie moléculaire, que sur les aspects microbiologiques, ainsi qu'à Lucile Jocteur Montrozier pour toutes les réponses apportées quant à mes questionnements sur le sol, et sa contribution dans les corrections.*

*Je voudrais également remercier LibraGen S.A., société avec laquelle j'ai eu la chance de pouvoir travailler. Je suis notamment très reconnaissante envers Renaud Nalin, et Benjamin Gilet, grâce à qui j'ai fait mes premiers pas dans la recherche. Les collaborations qui s'en sont suivies, notamment dans ce travail de thèse, furent très enrichissantes.*

*Merci également à Laurent Philippot, Ivan Matic, et Gisèle Laguerre pour leurs conseils avisés, notamment pour l'avancement du projet sur les axes traitant sur la recombinaison.*

*Je n'oublie pas tous les autres membres de l'ex équipe 4 qui m'ont chaleureusement accueillie et éclairée dans mon travail de recherche : merci notamment à Tim, Saliou, Pascal, Hélène, Hasna, Aude, Sandrine, Geneviève, Isabelle, Xavier, Catherine, Dominique, Jann, Denis, Alessandra, David, Franck et Van.*

*Ces quelques dernières notes sont pour toutes les personnes qui m'ont soutenues et encouragées pendant cette thèse, qui ont été d'une manière ou d'une autre une escale bienvenue dans les méandres des constructions génétiques toxiques...*

*Merci Benoît, pour avoir toujours été là dans les meilleurs moments comme dans les plus sombres, merci pour ta disponibilité, pour tous les services donnés, et pour m'avoir écouté (bien que je ne sois pas toujours compréhensible et légèrement soporifique !), ton amitié fut un vrai soutien. Merci à toi aussi Tiphaine ! Ton franc parler et tes pointes ironiques en ont surpris plus d'un, mais ça fait partie de ton charme ! J'ai passé d'excellentes années en votre compagnie. Merci pour tout !!! Merci pour les parties de fous rires, pour toutes ces petites soirées tarots, et toutes ces sorties champêtres : inoubliables ! D'ailleurs, un balbuzard vaut mieux que deux falconiformes ;)...*

*Merci à Servane, Julien, Mathieu, Sébastien, Aurélie, Romain, Nico, Carine, Elodie, Chloé, Marie-Laure, Charline, Nadège, mes compagnons de galère lyonnais (par vous, avec vous, mais pas sans vous ! je resterais lyonnaise dans l'âme). Une mention particulière pour Servane, notre fée à tous !!! Je te dois beaucoup...*

*Merci Dominique pour toutes tes charmantes et nombreuses attentions que je note et garde en mémoire, merci à toi Ghislaine pour tout ce que tu fais pour nous, qui contribue à mettre beaucoup de gaîté dans le labo, merci à Edwige pour ce que tu es et comme tu es ! Ne change surtout pas ! Merci à Jann, à Denis, et à Dad, mes compagnons d'écriture !!! Merci à Emilie, Martina, Steph, mes partenaires pour la danse et le roller !!! Merci à Maryline et Jean, amis d'escalades ! Merci à Catherine, Alessandra, Delphine, Hasna, Joel, Elodie, Dominique, Aude, David et Jérémy. Merci à Maude, Céline, Amy, et étudiants d'ECL. Merci à Bénédicte, Emelyne, Eléonore, Olivier, Fred, Karima, et Xavier. Merci pour toutes les marques d'attention et d'encouragements de chacun...*

*Une petite dédicace toute spéciale à Andrew Bird qui sans le savoir a été d'une très agréable compagnie (et source d'inspiration ? d'apaisement ?) pendant cette période rédactionnelle.*

*Pour finir (last but not least), je voudrais remercier ma famille. Merci d'avoir cru en moi, d'avoir énormément fait pour moi, de m'avoir poussée au delà de mes inquiétudes, et d'avoir toujours été là, tout simplement!*

*...Merci !*



# SOMMAIRE

INTRODUCTION GENERALE .....	17
-----------------------------	----

CHAPITRE I : Synthèse bibliographique .....	24
---------------------------------------------	----

Des bactéries du sol à l'exploitation de leur ADN: l'approche métagénomique

1	Introduction au système sol .....	25
1.1	La fraction minérale.....	26
1.2	La structure du sol .....	28
1.3	Porosité, eau, et oxygénation.....	30
1.4	Présence de microenvironnements dans le sol .....	32
2	Les bactéries dans l'environnement sol.....	33
2.1	Distribution des bactéries dans le sol .....	33
2.2	Les propriétés de la surface bactérienne.....	34
2.3	Accès aux cellules bactériennes telluriques.....	37
3	Les molécules ADN dans la matrice sol.....	38
3.1	L'ADN extracellulaire dans le sol .....	39
3.2	Interaction entre ADN et composants du sol.....	40
4	Extraction du métagénome .....	41
4.1	Isolement de l'ADN après lyse cellulaire <i>in situ</i> .....	41
4.2	Isolement de l'ADN après lyse cellulaire <i>ex situ</i> .....	43
5	Exploitation du métagénome .....	45
5.1	Analyse de la communauté bactérienne via l'outil PCR.....	46
5.2	Séquençage directe de l'ADN métagénomique.....	47
5.3	Construction et exploitation des banques métagénomiques .....	48
6	Conclusion .....	50

CHAPITRE II : Résultats .....	52
Facteurs influençant la qualité de l'ADN bacterien extrait du sol	

## High molecular weight dna recovery from soils prerequisite for biotechnological metagenomic library construction

1	Introduction .....	54
2	Materials and Methods .....	55
2.1	Soil samples.....	55
2.2	Extraction of soil bacteria.....	55
2.3	Bacterial counts .....	56
2.4	Enrichment cultures .....	56
2.5	Cell lysis and DNA purification .....	57
2.6	DNA fragment size determination by pulsed field electrophoresis.....	57
2.7	Direct extraction and purification of DNA.....	57
2.8	Structure of soil microbial community determined by ribosomal intergenic spacer analysis (RISA) .....	57
3	Results and Discussion .....	58
3.1	Extraction and purification of soil DNA .....	58
3.2	DNA fragment size.....	60
3.3	Enrichment of soil bacteria.....	61
3.4	Structure of the soil bacterial community.....	62
4	Acknowledgments .....	63
5	References .....	63

## Conclusions et Perspectives



CHAPITRE III : Résultats .....	71
Développement d'une nouvelle approche metagenomique : capture <i>in situ</i> de l'ADN métagénomique	

Etude et développement de systèmes de régulation pour le contrôle de l'expression de gènes toxiques	
1	Introduction ..... 75
2	Matériels et Méthodes ..... 78
2.1	Bactéries et plasmides ..... 78
2.2	Digestions enzymatiques ..... 79
2.3	Amplification de l'ADN par PCR ..... 79
2.4	Clonages ..... 80
2.5	Préparation des cellules pour l'électroporation ..... 81
2.6	Transformation ..... 81
2.7	Modification chromosomique..... 82
2.8	Criblage phénotypique..... 82
2.9	Test d'efficacité ..... 84
2.10	Cinétiques de croissance..... 84
3	Résultats et Discussion ..... 84
3.1	Contrôles et optimisations avant clonage de gènes toxiques..... 84
3.1.1	Transformabilité des cellules électrocompétentes :..... 84
3.1.2	Optimisations des techniques de clonage :..... 85
3.1.3	Vérification d'une expression inductible..... 86
3.2	Clonage des gènes toxiques ..... 87
3.2.1	La transformation normalisée, indicateur de la réussite du clonage de gènes toxiques ..... 87
3.2.2	Tests de différents promoteurs pour le contrôle de l'expression du gène <i>relF</i> ..... 89
3.2.3	Développement d'un nouveau système de régulation pour le contrôle de l'expression du gène <i>colE3</i> ..... 91
4	Conclusion ..... 96

## Développement d'un outil de contre sélection positive pour piéger l'ADN de la microflore tellurique

1	Introduction .....	98
2	Matériels et Méthodes .....	100
2.1	Bactéries et plasmides .....	100
2.2	Manipulations enzymatiques de l'ADN .....	101
2.3	Construction du plasmide pPRE.....	101
2.4	Modifications chromosomiques de <i>E.coli</i> .....	102
2.5	Test d'efficacité des constructions toxiques.....	102
2.6	Extraction de l'ADN métagénomique .....	103
2.7	Transformation / recombinaison ADN dans la souche réceptrice .....	103
2.8	Analyses de séquences.....	104
3	Résultats et Discussion .....	104
3.1	Définition et modifications de la souche réceptrice : .....	104
3.2	Définition du vecteur pour le système de contre sélection positive .....	105
3.3	Définition des zones de recombinaison .....	106
3.4	Construction du système de contre sélection positive .....	107
3.5	Efficacité du système de contre sélection à piéger l'ADN ciblé.....	108
3.6	Vérification d'un piégeage effectif de l'ADN par recombinaison .....	109
4	Conclusion.....	110

## Optimisation de l'efficacité de la cassette toxique du système de contre sélection positive pour la détection d'événements rares

1	Introduction .....	112
2	Matériels et Méthodes .....	113
2.1	Bactéries et plasmides .....	113
2.2	Manipulations enzymatiques de l'ADN .....	114
2.3	Test d'efficacité des constructions toxiques.....	114
3	Résultats et Discussion .....	115
3.1	Choix des gènes toxiques pour la construction de la cassette toxique .....	115
3.2	Contrôle de la souche réceptrice.....	116

3.3	Construction d'une cassette toxique inductible à deux gènes toxiques.....	117
3.4	Efficacité de la cassette toxique inductible.....	120
4	Conclusions .....	122

## Prospection des potentielles souches réceptrices pour accéder à l'ADN de la microflore tellurique

1	Introduction .....	123
2	Matériels et Méthodes .....	126
2.1	Extraction des cellules du sol et purification sur gradient de Nycodenz.....	126
2.2	Milieus, bactéries et plasmides .....	126
2.3	Préparation des cellules pour l'électrotransformation .....	127
2.4	Electrotransformation .....	127
2.5	Dénombrements des électrotransformants sur lames .....	128
2.6	Test d'efficacité des constructions toxiques.....	128
3	Résultats et Discussion .....	128
3.1	Développement d'une méthode pour la détection de souches électrotransformables cultivables .....	128
3.2	Efficacité de gènes toxiques dans différentes souches environnementales.....	131
4	Conclusion .....	132

## Conclusions et Perspectives

Conclusion générale .....	138
---------------------------	-----

Références bibliographiques.....	143
----------------------------------	-----

Annexes .....	162
---------------	-----

# LISTE DES TABLEAUX ET FIGURES

## Introduction

<u>Figure 1</u> : Approches conventionnelles et métagénomiques d'analyse des communautés microbiennes du sol.....	19
-------------------------------------------------------------------------------------------------------------------	----

## Chapitre I

<u>Figure 1</u> : Exemples d'unités structurales de minéraux.....	26
<u>Figure 2</u> : Altération des minéraux suivant le degré de solubilité des ions constitutifs .....	27
<u>Figure 3</u> : Triangle des textures d'un sol .....	27
<u>Figure 4</u> : Les agrégats dans le sol.....	30
<u>Figure 5</u> : Hydrophobicité de différentes espèces bactériennes .....	35
<u>Figure 6</u> : Représentation schématisée des deux voies d'extraction de l'ADN des bactéries contenues dans un échantillon de sol.....	41
<u>Tableau 1</u> : Caractéristiques de quelques phyllosilicates .....	28

## Chapitre II

<u>Figure 1</u> : Protocol of indirect bacterial DNA extraction and purification from soil .....	57
<u>Figure 2</u> : PFGE of bacterial soil DNA extracted by indirect method from Montrond soil and Châteaurenard soil .....	60
<u>Figure 3</u> : PFGE of bacterial soil DNA extracted by indirect method from Montrond and Châteaurenard soil after enrichment.....	61
<u>Figure 4</u> : Bacterial community structure of Montrond and Châteaurenard soils evaluated by RISA .....	62
<u>Figure 5</u> : Statistical analysis of data profiles performed by principal components analysis (PCA) for Montrond and Châteaurenard soils.....	63
<u>Table 1</u> : Châteaurenard and Montrond soil characteristics .....	55

## Chapitre III

<u>Figure 1</u> : Mécanisme de piégeage de l'ADN de la microflore tellurique dans une souche réceptrice .....	72
<u>Figure 2</u> : Principe de la contre sélection pour l'isolement des souches réceptrices transformées / recombinaisons .....	73
<u>Figure 3</u> : Nomenclature et points abordés pour la conception de la souche réceptrice.....	74

### Sous Chapitre III-1

<u>Figure 4</u> : Comparaison des profils de digestion entre les plasmides pMR32, pUCTc, et les plasmides extraits de transformants.....	85
<u>Figure 5</u> : Influence de la purification des plasmides par ultracentrifugation sur gradient de chlorure de césium sur les réactions de digestion.....	86
<u>Figure 6</u> : Contrôle de l'expression exercé par différentes constructions génétiques .....	87
<u>Figure 7</u> : Cinétique de croissance et vitesse de croissance comparée de <i>E. coli</i> contenant les plasmides pMMB190, pMR32, pMRTc20 et pFEB98 .....	90
<u>Figure 8</u> : Cinétique de croissance et vitesse de croissance comparée de souches <i>E. coli</i> TELI contenant les plasmides pBAD33 et pB33C3 .....	96
<u>Tableau 1</u> : Liste des souches <i>E. coli</i> utilisées et construites .....	78
<u>Tableau 2</u> : Liste et caractéristiques des plasmides utilisés et construits .....	79
<u>Tableau 3</u> : Liste des amorces, de leurs caractéristiques et des conditions d'amplification utilisées .....	80
<u>Tableau 4</u> : Fréquence de transformation de plasmides dans la souche <i>E. coli</i> JM109 en fonction des conditions d'utilisation des souches préparées électrocompétentes .....	85
<u>Tableau 5</u> : Caractéristiques de clonage et fréquences de transformation des constructions génétiques réalisées .....	88
<u>Tableau 6</u> : Efficacité des constructions toxiques réalisées.....	89

### **Sous chapitre III-2**

<u>Figure 9</u> : Emplacement des amorces Bad33bglIIIFw2 et Bad33pstIRv2 sur pBAD33 pour l'amplification de pB342 .....	106
<u>Figure 10</u> : Stratégie de construction du plasmide pPRE .....	107
<u>Tableau 7</u> : Liste des souches <i>E. coli</i> et plasmides utilisés ou construits .....	100
<u>Tableau 8</u> : Liste des amorces, de leurs caractéristiques et des conditions d'amplification utilisées .....	101
<u>Tableau 9</u> : Efficacité de pB33C3 en fonction de la quantité d'arabinose disponible pour l'induction de l'expression de <i>colE3</i> .....	105

### **Sous chapitre III-3**

<u>Figure 11</u> : Carte de restriction de pB33C3 .....	117
<u>Figure 12</u> : Stratégie de construction et de mobilisation de la cassette toxique .....	118
<u>Tableau 10</u> : Liste et caractéristiques des plasmides utilisés ou construits .....	113
<u>Tableau 11</u> : Liste des amorces, de leurs caractéristiques et des conditions d'amplification utilisées .....	114
<u>Tableau 12</u> : Taux d'échappement des gènes toxiques dans une construction génétique données .....	115

### **Sous chapitre III-4**

<u>Figure 13</u> : Intensité de fluorescence de <i>P. fluorescens</i> après acquisition de pSM1885 .....	130
<u>Tableau 13</u> : Comparaison des estimations des densités bactérienne d'une culture bactérienne suivant la méthode de dénombrement utilisée .....	129
<u>Tableau 14</u> : Efficacité de la cassette toxique de pFEB98 chez différentes espèces bactériennes .....	131
<u>Figure 14</u> : Profil de restriction de plasmides extraits de <i>A. Baylyi</i> (pFEB98) échappant à l'induction de l'expression toxique du gène <i>relF</i> .....	132

# LISTE DES ABREVIATIONS

(A)RISA	(Automated) Ribosomal Intergenic Spacer Analysis
A.nal	Acide nalidixique
AB	AntiBiotique
ADN	Acide DéoxyRiboNucléique
Amp	Ampicilline
Ara	Arabinose
ARN(m)	Acide RiboNucléique (messenger)
BAC	Bacterial Artificial Chromosome
BET	Bromure d'EThydium
BLAST	Basic Local Alignment Search Tool
BP	Bande Passante
Cm	Chloramphénicol
CTAB	CetylTriméthylAmmoniumBromide
DGGE	Denaturing Gradient Gel Electrophoresis
dNTP	déoxy Nucléotide Tri Phosphate
DO	Densité Optique
DVLO	Derjaguin-Landau-Verwey-Overbeck
EDTA	Ethyl Diamine Tétra Acétate
EPS	ExoPolySaccharide
GFP	Green Fluorescent Protein
GM	Génétiquement Modifiée
IPTG	IsoPropyl $\beta$ -D-ThioGalactopyranoside
IR	Infra Rouge
Kb	Kilobase
Kn	Kanamycine
LB	Luria Bertani
Leu	Leucine
LPS	LipoPolySaccharide
MEPS	Minimal Efficient Processing Segment
NCBI	National Center for Biotechnological Information

ORF	Open Reading Frame
Pb	Paire de base
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PEG	PolyEthylèneGlycol
PVPP	PolyVinylPyrrolidone
PZC	Point Zéro Charge
QSP	Quantité Suffisante Pour
RBS	Ribosome Binding Site
SAP	Shrimp Alcaline Phosphatase
SDS	Sodium Dodecyl Sulfate
SIP	Stable Isotope Probing
Sm	Streptomycine
Sp	Spectinomycine
T	Transformation
TBE	Tris Borate EDTA
Tc	Tétracycline
Tm	Temperature melting
TN	Transformation Normalisée
T-RFLP	Terminal Restriction Fragment Length Polymorphism
Trp	Tryptophane
Ts	Thermosensible
TSA	Tryptic Soy Agar
U	Unité
UFC	Unités Formant Colonies
USDA	United States Department of Agriculture
UV	Ultra Violet
X-Gal	5-bromo-4-chloro-3-indolyl- $\beta$ -D-Galactopyranoside





## INTRODUCTION GENERALE

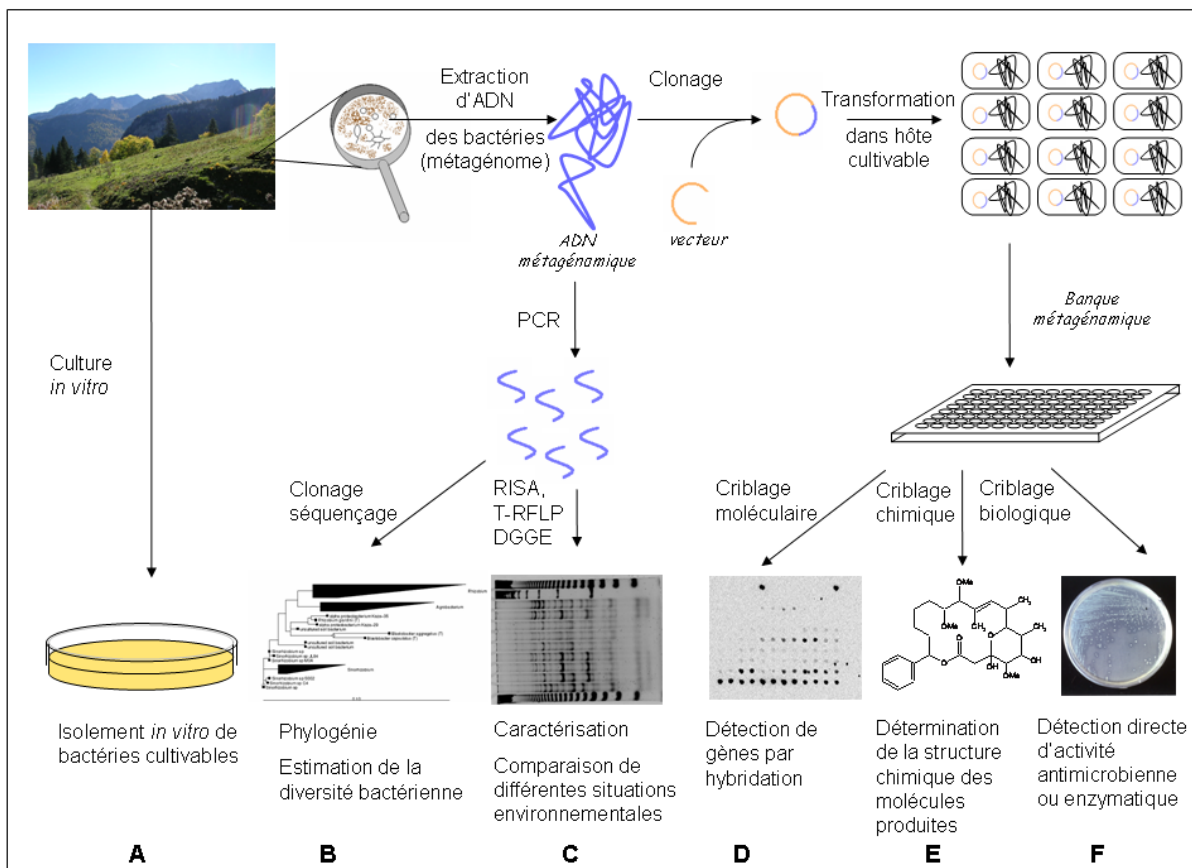
Les environs  $5 \times 10^{30}$  cellules bactériennes de notre planète représentent une biomasse totale quasi équivalente à celle des végétaux (Whitman *et al.*, 1998). Ces organismes microscopiques sont présents dans des environnements variés : des eaux douces aux eaux marines, des sols tropicaux aux zones désertiques, des sous sol profonds terrestres et océaniques (Whitman *et al.*, 1998) aux hautes couches atmosphériques comme la stratosphère (Imshenetsky *et al.*, 1978). Ils ont également colonisé des environnements dits « extrêmes » tels que les milieux hyper salins (Sorokin *et al.*, 2006), très acides (Tyson *et al.*, 2004), les sols glacés de l'antarctique (Vincent *et al.*, 2004) ou inversement la périphérie des sources hydrothermales en ébullition (Stetter, 1999). Les autres organismes de la planète n'ont pas échappé à leur main mise, l'homme ayant d'abord appris à les connaître du fait des épidémies causées par quelques bactéries pathogènes. Leurs effets bénéfiques sont également nombreux et l'homme les utilise depuis longtemps pour préparer ses aliments, certains procaryotes étant des probiotiques très efficaces (Heller, 2001). De nos jours, leurs propriétés sont exploitées dans les domaines agricoles et environnementaux pour améliorer la croissance de la plante (Lucy *et al.*, 2004), contrôler les ravageurs (Raaijmakers *et al.*, 2002), mais aussi en biorémediation où des bactéries interviennent dans la transformation et la dégradation de composés toxiques (Furukawa, 2003). Dans l'industrie pharmaceutique, les bactéries sont les producteurs de plus de 70 % des antibiotiques commercialisés à ce jour, ainsi que de nombreux anti-fongiques ou anti-prolifératifs (Demain, 1999).

Cependant, les bactéries de l'environnement sont encore très mal connues et le réservoir de ressources génétiques que constitue le monde microbien demeure largement sous exploité. La principale raison est la résistance de l'immense majorité (99 %) des bactéries à se multiplier *in vitro* (Amann *et al.*, 1995). L'arbre phylogénétique du vivant construit initialement par Woese (Woese, 1987) à partir des séquences des gènes ribosomiques et constamment adapté depuis à partir de nouvelles séquences témoigne de l'extrême diversité génétique au sein des bactéries. Parmi les 36 phyla procaryotiques caractérisés en 2002 par Hugenholtz (Hugenholtz, 2002) qui seraient 53 aujourd'hui (Schloss & Handelsman, 2004), 13 n'ont pour l'instant encore aucun représentant cultivé (Hugenholtz, 2002). Toute la

microbiologie fondamentale, médicale et industrielle depuis Pasteur s'est donc développée sur des isolats bactériens ne représentant qu'une infime proportion de la diversité microbienne.

Des efforts importants sont consacrés aujourd'hui afin d'améliorer les conditions du développement bactérien *in vitro*, en définissant par exemple des milieux de culture dont la composition physico-chimique est la plus proche possible de celle existant *in situ*. En utilisant des conditions oligotrophiques et en augmentant les temps d'incubation, Janssen et ses collaborateurs (Janssen *et al.*, 2002) ont ainsi pu isoler des représentants des phylums *Verrucomicrobia* et *Acidobacteria*. D'autres stratégies reposent sur des protocoles particuliers comme celui qui consiste à encapsuler les cellules extraites d'un environnement donné dans des petits volumes d'agar puis de proposer aux cellules isolées un large éventail de conditions de culture différentes. Cette méthode présente l'avantage d'isoler physiquement chaque cellule bactérienne, tout en permettant les échanges moléculaires entre bactéries, grâce à la porosité de l'agar. Zengler et ses collaborateurs (Zengler *et al.*, 2002) ont ainsi pu constater le développement de microcolonies de bactéries jusqu'à présent non cultivées. Cette stratégie semble efficace pour des organismes appartenant à différents phyla bactériens colonisant des environnements marins et terrestres, mais la détermination des conditions de croissance des microorganismes *in vitro* reste longue, fastidieuse et très souvent infructueuse. Certaines bactéries sont sensibles à des composés du milieu, d'autres nécessitent pour leur croissance une combinaison particulière de conditions environnementales ou des facteurs de croissance très spécifiques (Ferrer *et al.*, 2003; Gabor *et al.*, 2004b). Malgré quelques avancées, la proportion de bactéries cultivées *in vitro* ne s'accroît que très lentement.

Ces biais dans la propagation en culture pure de la plupart des bactéries ont conduit à privilégier des approches s'affranchissant du développement *in vitro*. C'est ainsi que depuis le début des années 1990 la microbiologie environnementale s'est dotée d'un nouveau concept en restreignant l'étude des bactéries à leur ADN directement extrait de l'environnement (**Figure 1**). Des protocoles ont été développés afin d'extraire et de purifier l'ADN à partir d'échantillons environnementaux. L'extrait obtenu contient potentiellement l'ensemble des génomes microbiens présents dans l'échantillon et qui de ce fait a été appelé le métagénome (Rondon *et al.*, 2000). La première exploitation des molécules d'ADN extraites et purifiées des milieux naturels a été portée sur l'estimation de la diversité des bactéries (Amann *et al.*, 1995; Hugenholtz *et al.*, 1998; Torsvik *et al.*, 1990). Très rapidement les bases de données se sont enrichies de nombreuses séquences venant de bactéries non cultivées permettant de



**Figure 1 : Approches conventionnelles et métagénomiques d'analyse des communautés microbiennes du sol (Lombard *et al.*, 2006)**

Les microorganismes d'un sol peuvent être étudiés par une approche culturale (A), cependant moins de 1% seulement se développent sur les milieux classiquement utilisés. Une alternative consiste extraire puis analyser l'ADN des bactéries présentes dans le sol étudié. L'ADN métagénomique obtenu est soit analysé après amplification ciblée sur des gènes à critères taxonomiques ou des gènes codant une fonction écologique donnée (B,C), soit d'abord cloné dans un hôte cultivable avant analyse (D, E, F).

PCR : Polymerase Chain Reaction, RISA : Ribosomal Intergenic Spacer Analysis, T-RFLP : Terminal Restriction Fragment Length Polymorphism, DGGE : Denaturing Gradient Gel Electrophoresis

positionner ces organismes inconnus dans les arbres phylogénétiques existants. Par comparaison avec les isolats obtenus sur milieux de culture, cette approche a permis de réévaluer les niveaux de la diversité bactérienne au sein d'un écosystème. Cette approche s'est également déclinée selon toute une panoplie de variantes méthodologiques pour caractériser très rapidement une communauté complexe. Ces méthodologies (**Figure 1**) basées sur la migration électrophorétique des produits PCR (*polymerase chain reaction*) produisent des profils dans lesquels chaque bande d'ADN peut être associée plus ou moins spécifiquement à une unité taxonomique opérationnelle (Kirk *et al.*, 2004; Ranjard *et al.*, 2000). Applicables aux gènes ribosomiques comme à d'autres gènes cibles, ces techniques ont notamment été utilisées pour comparer entre elles différentes situations environnementales (Gros *et al.*, 2006; Muller *et al.*, 2002; Mummey & Stahl, 2004).

Une variante méthodologique d'exploitation du métagénome consiste à cloner l'ADN extrait de l'environnement dans des cellules bactériennes cultivables, généralement *Escherichia coli* (Handelsman, 2004) (**Figure 1**). Les banques d'ADN métagénomique qui en résultent sont ainsi constituées de clones différenciés contenant chacun un fragment d'ADN environnemental intégré dans un vecteur adapté et multiplié dans un hôte bactérien domestiqué comme peut l'être *E. coli*. Cette approche présente de nombreux avantages. Elle permet l'affranchissement de l'étape PCR qui présente des biais liés à l'amplification préférentielle de certains types d'ADN au détriment d'autres (von Wintzingerode *et al.*, 1997). De plus, cette approche ouvre des perspectives élargies pour l'exploitation du métagénome grâce à la possible expression des gènes clonés dans leur nouvel hôte. Ceci accroît les capacités d'analyse des génomes bactériens non cultivés avec des applications industrielles potentielles pour détecter de nouvelles enzymes ou de nouveaux composés bioactifs (Daniel, 2005). A ces fins tant fondamentales qu'appliquées des banques métagénomiques ont été construites à partir d'ADN extrait d'échantillons de sol (Courtois *et al.*, 2003; Liles *et al.*, 2003; Rondon *et al.*, 2000), d'eau de mer (Beja *et al.*, 2000; Stein *et al.*, 1996), ou encore de fèces humain (Breitbart *et al.*, 2003).

Un recouvrement et une exploitation exhaustifs de la diversité bactérienne, très importante dans certains environnements, nécessitent cependant la construction, la manipulation et le criblage de banques métagénomiques en conséquence. Par exemple, la diversité bactérienne a été estimée dans certains sols à  $10^9$  cellules par gramme de sol différenciées en environ 10 000 espèces génomiques dominantes différentes (Torsvik *et al.*,

2002). Une telle diversité nécessiterait la création de banques de plusieurs millions de clones de manière à ce que le génome d'un représentant de chacune des espèces répertoriées y soit représenté, qu'elles soient abondantes dans l'environnement ou à effectif réduit (Curtis & Sloan, 2004; Riesenfeld *et al.*, 2004b). Les possibilités techniques actuelles de la plupart des laboratoires (au moins dans le domaine académique) ne permettent pas de répondre à de telles attentes (Daniel, 2005). En dépit de ces limitations technologiques, la quantité d'informations génétiques contenues dans les banques d'ADN métagénomique décrites dans la littérature est déjà très importante et ne peut être exploitée sans des méthodes de criblage adaptées, notamment lorsque des gènes peu fréquents sont recherchés parmi la microflore.

A ces limitations d'ordre logistique s'ajoutent celles liées aux caractéristiques physico-chimiques des environnements dans lesquels se développent les bactéries, en particulier le sol, qui vont biaiser plus ou moins fortement la récupération de l'ADN bactérien et sa purification. Nous verrons plus loin que la colonisation microbienne hétérogène de la matrice complexe qu'est le sol (Ranjard & Richaume, 2001) s'accompagne d'un taux de recouvrement des cellules bactériennes (et de leur ADN) très disparate et variable selon les sols. L'efficacité des traitements d'extraction est également fortement conditionnée par la composition chimique du sol, l'adsorption des cellules et de leur ADN sur certains constituants comme les particules d'argiles (Bakken, 1985; Frostegard *et al.*, 1999). Enfin des problèmes de purification de l'ADN peuvent survenir avec des conséquences nuisibles sur l'efficacité de clonage (Gabor *et al.*, 2003).

A l'initiation de ce travail de thèse en septembre 2003, les approches microbiologiques s'appuyant sur l'exploitation du métagénome bactérien apparaissaient extrêmement prometteuses pour contourner les biais liés aux problèmes de culture *in vitro* de la plupart des bactéries. Le séquençage des gènes ribosomiques détectés dans le métagénome avait déjà permis d'accroître significativement les connaissances sur le niveau (très élevé) de la biodiversité bactérienne. Plusieurs articles scientifiques présentaient la construction de banques d'ADN environnemental riches de dizaines de milliers de clones recombinants, dont les applications intéressaient aussi bien la recherche académique, fondamentale, que la recherche industrielle. Ces nouveaux concepts pouvaient être exploités dans divers domaines dont l'agro-alimentaire et la pharmacie. Cette époque a vu le développement de plusieurs jeunes sociétés dont la compagnie française LibraGen S.A. qui a pour vocation de valoriser les ressources génétiques inexplorées et étendues que représente la microflore non cultivée.

Une autre analyse du développement de ces techniques mettait en évidence leurs limitations tant technologiques que techniques pour les raisons que nous avons rapidement indiquées ci-dessus. Se posaient alors des questions sur les possibilités pour améliorer la technologie existante ou en développer de nouvelles, de façon à limiter les biais, accéder à une fraction plus importante de la communauté bactérienne, et avoir une meilleure représentativité de la diversité. Pouvait-on trouver les moyens d'accéder à l'information génétique de populations bactériennes auxquelles ne donnaient pas encore accès les techniques métagénomiques développées ? Était-il possible de contourner les biais liés à la lyse, à l'extraction cellulaire et aux opérations de clonage ?

Nos objectifs au cours de cette thèse ont été de tenter de répondre à ces questions, tant en cherchant à améliorer les protocoles existants, qu'en développant des approches novatrices pour lesquelles ne subsisterait plus aucun biais d'extraction et de purification de l'ADN ou de clonage. Nous avons toutefois conscience qu'apparaîtraient nécessairement d'autres types de limitations. Nous avons donc d'abord recherché comment améliorer la représentativité de la diversité bactérienne dans les banques d'ADN métagénomique extrait du sol en tentant d'augmenter la taille des inserts, ce qui aurait également pour conséquence de diminuer le nombre de clones à manipuler pour un taux de couverture identique. Les résultats (voir chapitre II) ont apporté de nouvelles informations sur le mode de vie des bactéries dans le sol, la façon dont était économisée l'énergie pendant les périodes de dormance et l'efficacité des mécanismes de réparation de l'ADN à la reprise d'un métabolisme plus actif. Ces travaux ont aussi montré les limites que peuvent rencontrer les protocoles d'extraction de l'ADN à partir du sol, ce qui a motivé notre choix de développer des approches alternatives pour s'affranchir des étapes d'extraction, de lyse et de purification de l'ADN extrait du sol. La stratégie que nous avons privilégiée est basée sur l'utilisation d'une souche bactérienne susceptible d'acquérir *in situ* (sol) des gènes de la microflore indigène. Cette stratégie n'est toutefois pertinente que si elle inclut des artifices moléculaires permettant de sélectionner positivement les clones recombinants afin de pouvoir les exploiter ultérieurement (voir chapitre III).

Les voies d'accès à l'ADN métagénomique pour son analyse ultérieure sont ainsi présentées dans ce mémoire en trois chapitres :

Le chapitre I est une synthèse bibliographique sur l'extraction de l'ADN des bactéries présentes dans le sol, et les moyens développés pour leur analyse et exploitation. Une attention particulière est portée sur l'environnement sol, sur ses particularités physico-chimiques et leurs impacts sur l'extraction du métagénome, et par voie de conséquence sur les études métagénomiques réalisées à partir de l'information génétique extraite.

Le chapitre II expose les résultats obtenus concernant les facteurs affectant la qualité de l'extraction. L'objectif de cette étude était de définir les moyens pouvant être développés pour augmenter la taille des fragments d'ADN extraits de l'environnement sol.

Le chapitre III présente la nouvelle approche métagénomique proposée dans cette thèse. Cette approche totalement innovante pour l'étude de la diversité de séquences des gènes ciblés dans l'environnement sol a nécessité d'importants développements technologiques et optimisations afin de valider son application. Elle a notamment nécessité le développement d'un système de contre sélection positive efficace permettant la seule croissance des souches ayant piégé l'information génétique ciblée. L'ensemble des développements effectués est présenté en quatre sous chapitres portant sur (1) l'optimisation du contrôle d'une expression toxique pour la cellule hôte (2) les premiers essais expérimentaux avec le système développé (3) les optimisations nécessaires pour son application (4) une ouverture sur le choix d'une souche réceptrice pour capturer l'information génétique de la microflore tellurique.

Une conclusion générale et les perspectives d'application des travaux développés ici clôtureront ce manuscrit.





# **CHAPITRE I : SYNTHESE BIBLIOGRAPHIQUE**

## **DES BACTERIES DU SOL A L'EXPLOITATION DE LEUR ADN: L'APPROCHE METAGENOMIQUE**

Les problèmes liés à la culture des bactéries *in vitro* ont conduit à l'utilisation d'approches contournant l'isolement cellulaire pour étudier la diversité bactérienne d'un environnement donné. Ces approches nécessitent l'extraction préalable de l'ADN des bactéries présentes dans l'échantillon étudié. L'extraction de cette information génétique (métagénome) à partir de la matrice sol est particulièrement délicate, les propriétés physico-(bio)-chimiques du sol affectant plus ou moins fortement l'accès au métagénome.

Pour résoudre ces problèmes d'accessibilité, les relations qu'entretiennent les bactéries telluriques avec leur environnement doivent être prises en compte. La structuration de la matrice sol, la distribution des bactéries dans cette matrice, l'interaction des bactéries ou de l'ADN lui-même avec les particules de sol, sont autant de paramètres à considérer pour réaliser dans les meilleures conditions possibles l'extraction du métagénome d'un échantillon de sol.

A ce jour, deux stratégies d'extraction du métagénome ont été développées. Dans la première, dite directe, les bactéries sont lysées *in situ*. L'ADN alors libéré est extrait de la matrice puis purifié. La seconde stratégie vise à séparer les cellules bactériennes des autres constituants de la matrice avant d'employer les traitements lytiques et de purification de l'ADN. Avantages et inconvénients de chacune des stratégies ont été rapportés dans la littérature. L'utilisation de l'une ou l'autre méthode est souvent dépendante de l'exploitation subséquente de l'ADN métagénomique qui en est faite.

L'objectif de ce chapitre bibliographique est de souligner l'impact des propriétés physico-chimiques du sol sur l'extraction du métagénome. En effet, la complexité et l'hétérogénéité de cette matrice conditionnent d'une part la localisation des bactéries dans le sol, d'autre part leur adsorption, comme celle de leur ADN, sur les constituants du sol. Comprendre les différentes interactions entre composantes biotiques et abiotiques requiert la connaissance du système sol et de ses propriétés physico-chimiques.

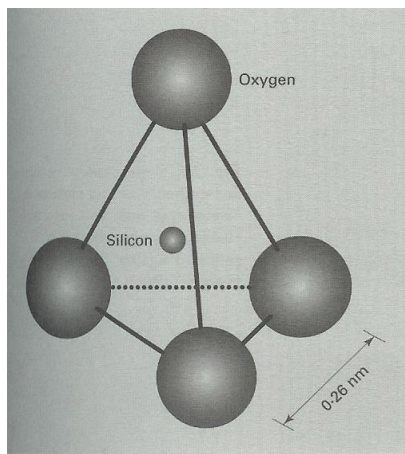
Nous proposons donc d'aborder le système sol dans une première partie de cette synthèse bibliographique, selon un point de vue tourné vers la physico-chimie. Les processus de formation et d'évolution générale des sols ne seront pas détaillés, la couverture végétale et les systèmes rhizosphériques ne seront que rapidement évoqués, seuls les points clés pour la compréhension des interactions entre cellules bactériennes (et leur ADN) et les composantes du sol seront précisés. Dans une deuxième et troisième partie, les interactions des bactéries et de leur ADN avec la matrice sol seront approfondies. De ces données fondamentales pourront être déduits les principes ayant prévalu aux développements technologiques d'extraction du métagénome, présentés dans une quatrième partie. Les différentes exploitations tant fondamentales qu'appliquées qui peuvent en être faites seront évaluées dans la dernière partie.

## **1 INTRODUCTION AU SYSTEME SOL**

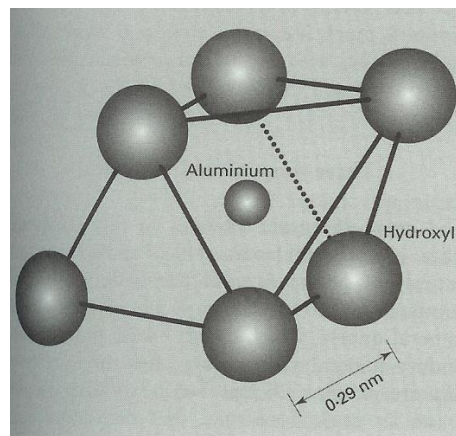
Le sol est à l'interface entre la lithosphère (roche mère sous jacente), l'atmosphère et l'hydrosphère (Alexander, 1977) et sert de support à une partie de la biosphère (Gobat *et al.*, 2003). Il est constitué de cinq composantes majeures : fraction minérale, matière organique, eau, air, et organismes vivants (Alexander, 1977). Les matières organiques et minérales s'organisent en structures à l'origine de la création de vides, dont les volumes ainsi formés (ou porosité) sont occupés par l'air et l'eau. Cette structuration résulte des propriétés physico-chimiques propres à ces deux fractions.

Bien que la fraction organique joue un rôle non négligeable dans ces interactions, une attention plus particulière est ici portée sur la fraction minérale et notamment sur les silicates et phyllosilicates (argiles) qui en sont les constituants majeurs. Les argiles présentent en effet une grande réactivité et sont fortement impliqués dans des interactions avec d'autres composantes du sol, dont les cellules bactériennes et leur ADN. Pour expliquer leurs propriétés physico-chimiques, des précisions sont apportées sur les roches mères et leurs minéraux constitutifs, les processus de désagrégation physique et d'altération à l'origine de la fraction minérale résultante dans les sols.

La structure d'un sol ayant des répercussions sur la localisation des bactéries, ce point sera également abordé via le processus de formation des agrégats, ainsi que les conséquences sur l'occupation de l'eau et l'air dans les porosités formées. L'ensemble de ces données éclairera les aspects de distribution des microorganismes dans le sol.



(A)



(B)

**Figure 1 : Exemples d'unités structurales de minéraux** (White, 2006)

(A) Unité structurale de silice (B) Unité structurale d'alumine (distances inter-atomiques non respectées)

## 1.1 La fraction minérale

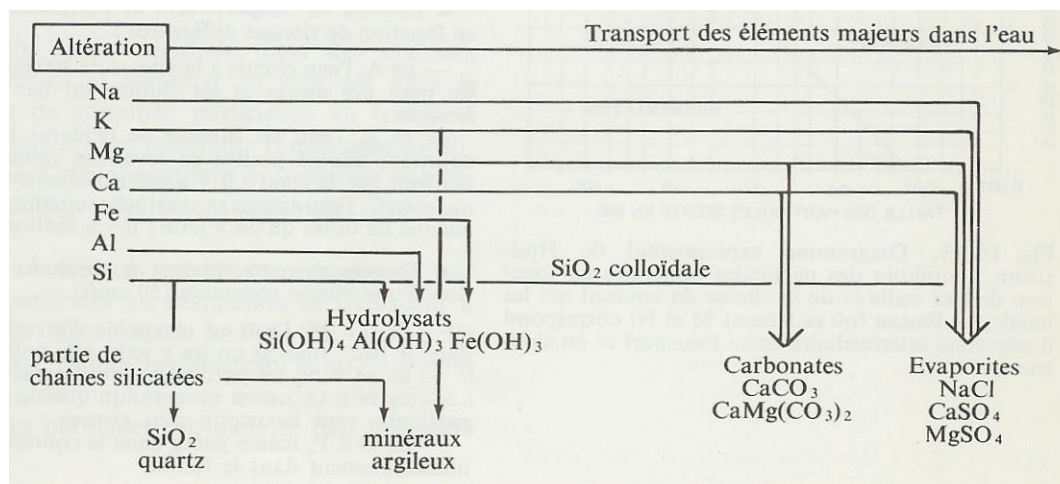
La fraction minérale dépend du type de roche mère (ou roche primaire) à partir duquel le sol s'est formé. Elle est en effet le résultat de la désagrégation et de l'altération de la roche mère. Trois types de roche sont distinguées selon des critères géologiques et géochimiques (Best, 1995; Tyler, 2004) : **Les roches magmatiques** (également appelé ignées) se forment suite au refroidissement du magma, quelle que soit la profondeur (dans la lithosphère ou en surface) et donc la pression considérée. **Les roches sédimentaires** proviennent du dépôt puis consolidation de particules ou de matériaux dissous, tous deux issus de l'altération de la roche mère et du transport dans les zones de sédimentation. **Les roches métamorphiques** résultent de la recristallisation en phase solide des roches magmatiques et sédimentaires (en général en présence d'un fluide aqueux). Ce phénomène, qui a lieu pour des conditions de température élevées (mais submagmatiques) et des conditions de pression variables, entraîne une modification de la texture, structure et / ou composition de la roche originelle (Best, 1995). Les deux dernières catégories de roches dérivent donc de manière plus ou moins directe des roches magmatiques. Elles contiennent donc les minéraux formés dans les roches magmatiques ou les dérivés de ces minéraux.

### Les minéraux de la roche mère

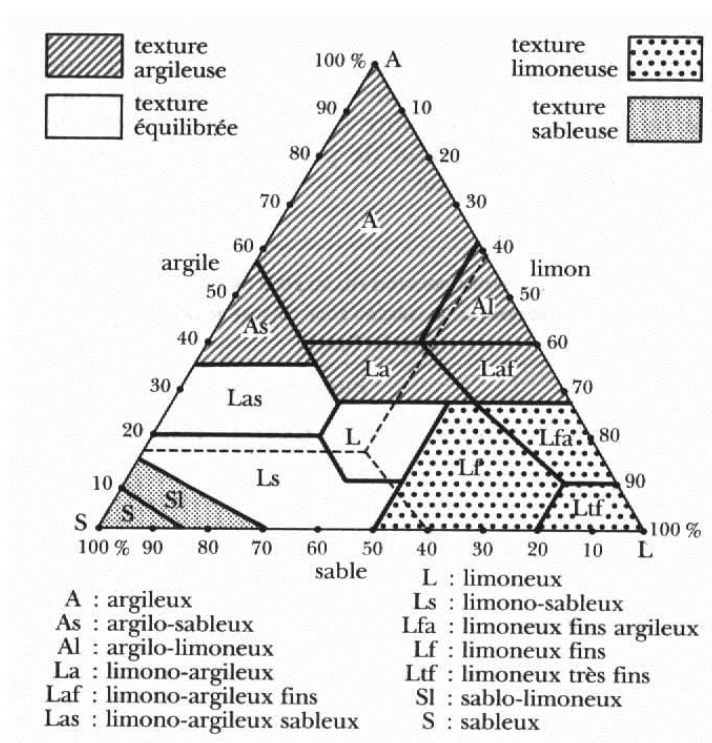
Le quartz est le minéral majoritaire des roches ignées formées en profondeur. Le cristal de quartz est composé de la polymérisation d'un atome de silice ( $\text{Si}^{4+}$ ) lié par covalence à 2 atomes d'oxygène ( $\text{O}^{2-}$ ). Le cristal de quartz  $(\text{SiO}_2)_n$  formé est souvent considéré comme inerte et inaltérable.

Les roches sont composées de nombreux autres minéraux silicatés. L'unité structurale majoritaire de ces minéraux est constituée de tétraèdres  $\text{SiO}_4^{4-}$ . D'autres unités structurales existent avec un atome d'aluminium  $\text{Al}^{3+}$ , ou de magnésium  $\text{Mg}^{2+}$  par exemple à la place de la silice. Ces éléments liés à des atomes d'oxygènes forment des structures tétraédriques, ou octaédriques suivant les cas (White, 2006) (**Figure 1**).

Ces unités structurales peuvent rester en structure cristalline simple (olivine), se lier en partageant des atomes d'oxygène pour former des chaînes (pyroxène), des associations de chaînes différentes en feuillets silicatés (ou phyllosilicates) (muscovite, biotite), ou en structures tridimensionnelles (feldspath). Les charges négatives restantes des minéraux silicatés sont soit neutralisées par des cations divers ( $\text{Al}^{3+}$ ,  $\text{Fe}^{3+}$ ,  $\text{Fe}^{2+}$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{K}^+$  et  $\text{Na}^+$ ), soit compensées par des arrangements structuraux internes (White, 2006).



**Figure 2 :** Altération des minéraux suivant le degré de solubilité des ions constitutifs (Dercourt & Paquet, 1981)



**Figure 3 :** Triangle des textures d'un sol (d'après l'USDA ou *United States department of agriculture*). La texture d'un sol est définie en fonction du pourcentage de sable, de limon et d'argile qui le composent.

### Processus de formation de la fraction minérale

La fraction minérale des sols résulte de l'action de deux principaux phénomènes :

**La désagrégation physique** au cours de laquelle les alternances chaud / froid, le gel, le vent ou l'eau fractionnent la roche en morceaux de plus en plus petits. Les particules obtenues conservent la composition minéralogique initiale (Gobat *et al.*, 2003).

**L'altération** faisant intervenir l'eau, associée ou non à l'oxygène, au gaz carbonique CO<sub>2</sub> ou à des acides organiques (Gobat *et al.*, 2003). Elle provoque une transformation des minéraux primaires avec formation de minéraux secondaires (notamment argiles) dont l'ensemble constitue le complexe d'altération (Duchaufour, 1977). Elle se manifeste selon plusieurs voies dont la principale est l'hydrolyse (neutre ou acide).

Cette voie d'altération majeure des roches cristallines conduit à la destruction partielle des chaînes et leurs réorganisations en feuillets silicatés ou minéraux argileux, voire une destruction complète des chaînes et à la formation d'hydrolysats, suivant les minéraux et le degré de solubilité des ions constitutifs (Dercourt & Paquet, 1981) (**Figure 2**). Notamment, les roches riches en silice (et donc acides) sont plus résistantes à l'altération que les roches pauvres en silice (basiques) mais riches en fer et magnésium.

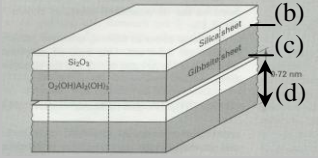
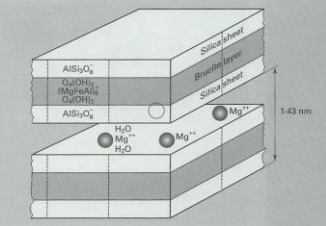
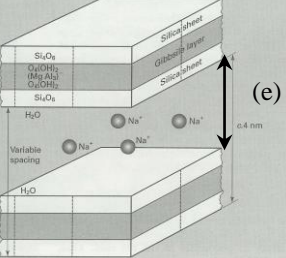
Ainsi le quartz SiO<sub>2</sub> est résistant à l'altération chimique et mécanique, les feldspaths s'altèrent d'autant plus vite qu'ils sont pauvres en silice, et la biotite s'altère rapidement par la perte de fer et d'ions interfoliaires.

### La fraction minérale : importance des argiles dans les propriétés physico-chimiques des sols

La fraction minérale contenue dans le sol est ainsi issue des processus de désagrégation et altération décrits ci-dessus. Elle est composée de particules de différentes tailles, généralement classées selon leur diamètre. Par ordre de taille décroissante sont distingués la fraction grossière supérieure à 2 mm (bloc, cailloux, etc.), de la terre fine inférieure à 2 mm (sables, limons, et argiles). La proportion relative des éléments de la terre fine détermine la texture d'un sol (**Figure 3**), propriété conditionnant directement la structure du sol (Gobat *et al.*, 2003), c'est-à-dire l'organisation des particules et des vides.

Les argiles sont plus particulièrement le résultat de l'altération des roches par hydrolyse des minéraux silicatés. Les minéraux argileux formés sont des phyllosilicates, constitués de feuillets organisés en couche. Les argiles sont classées suivant le nombre de feuillets qui constituent une couche et l'épaisseur constante ou variable entre les couches (**Tableau 1**). Certaines argiles (argiles gonflantes) sont en effet capables d'expansion suite à l'hydratation des couches interfoliaires (Gobat *et al.*, 2003).

**Tableau 1 : Caractéristiques de quelques phyllosilicates (D'après White, 2006)**

Argile minéralogique	Ratio Si :Al (par couche)	Surface spécifique	CEC <sup>a</sup> (cmol charge+.kg <sup>-1</sup> )
<b>Kaolinite</b>	1 :1 	05-40	05-25
<b>Illite</b>	2 :1	100-200	20-40
<b>Vermiculite</b>	2 :1 	300-500	150-160
<b>Montmorillonite (saturé Na) gonflante</b>	2 :1 	750	80-150

<sup>a</sup> Capacité d'échange cationique

(b) feuillet silicaté

(c) feuillet d'alumine

(d) couche phyllosilicate

(e) distance inter couche



Une autre composante de la réactivité des particules argileuses est leur surface spécifique. Plus leur surface disponible est élevée, plus leur réactivité sera prédominante parmi les éléments de la texture (Alexander, 1977; White, 2006). Notamment, la capacité d'adsorption de molécules et ions à la surface des argiles augmente avec leur surface disponible par unité de masse (White, 2006) (**Tableau 1**).

Par ailleurs, les phyllosilicates argileux sont des colloïdes possédant d'autres propriétés qui vont influencer la structure, la porosité ou la capacité d'échange (cation)-ionique du sol. En effet, les argiles sont **électronégatives** en raison de la présence de valences négatives non satisfaites suite à des substitutions de cations à l'intérieur des couches (substitutions isomorphes). Les argiles sont **hydrophiles**, en particulier les argiles gonflantes, mais aussi lipophiles (elles sont utilisées pour absorber les excédents de corps gras). Enfin, les argiles présentent un **pouvoir de dispersion** suite aux forces de répulsion entre particules chargées négativement et **de floculation** lorsque la surface des particules est chargée de cations polyvalents ( $\text{Ca}^{2+}$  en milieu à pH neutre, et les ions alumineux  $\text{Al}^{3+}$  dans les milieux à pH acides). Ces propriétés varient également suivant les structures minéralogiques et les surfaces développées respectives (Gobat *et al.*, 2003) (**Tableau 1**) et régissent la formation d'agrégats, cimentent les particules et favorisent l'habitat microbien.

## 1.2 La structure du sol

### Interaction matière organique- minéraux

La matière organique, par ses propriétés physico-chimiques, interagit avec les particules de sol et ainsi participe à la structuration du sol. La matière organique fraîche provient des déchets des organismes, de leurs sécrétions ou des tissus morts. Elle est majoritairement héritée des débris végétaux.

La cellulose et les hémicelluloses, constituants des parois des cellules végétales, sont rapidement dégradés. Les polysaccharides sous-produits de leur dégradation, et d'autres macromolécules issues des débris végétaux, ont des propriétés d'agrégation essentielles à la structuration des sols (Chenu, 1993).

Autre élément du squelette végétal, la lignine, polymère complexe présentant des noyaux aromatiques, est plus résistante à la dégradation que les polysaccharides (Gobat *et al.*, 2003). Elle est avec les tannins à l'origine des matières organiques humifiées, qui constituent 70 à 80 % de la matière organique du sol (Parent & Velegol, 2004).

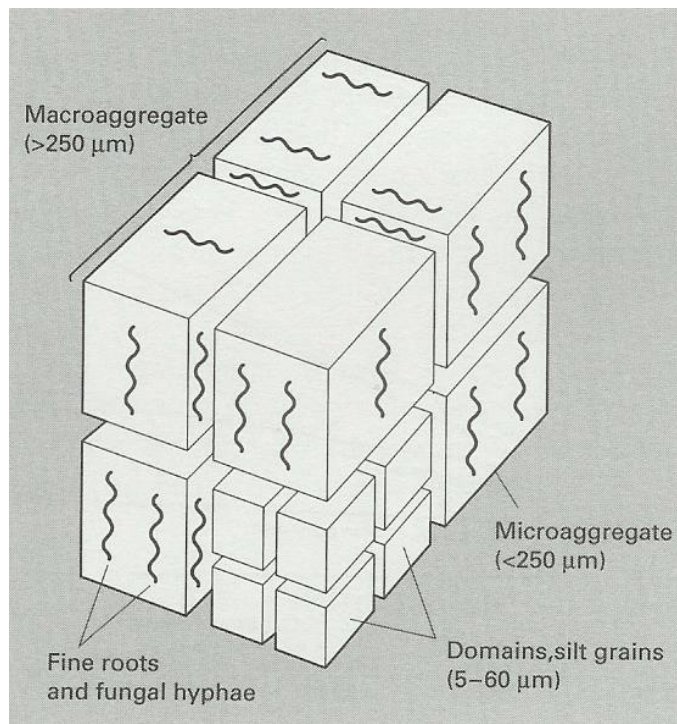
Ces composés organiques humifiées sont généralement des molécules aromatiques complexes, présentant divers degrés de polycondensation. Ils sont séparés en trois fractions suivant leur solubilité : les acides fulviques très solubles à tout pH, les acides humiques insolubles à pH acide, et l'humine insoluble. Selon le modèle de structure des composés humiques proposé par Andreux (Duchaufour, 1977), les acides fulviques sont composés d'un nucléus aromatique de petite taille et de longues chaînes aliphatiques ou peptidiques latérales (Gobat *et al.*, 2003). Ces chaînes présentent différents groupes fonctionnels tels que carboxyles, phénols, alcools, carbonyles, éthers et esters (Parent & Velegol, 2004). Les acides fulviques sont très réactifs par la présence de groupes carboxyle COOH qui leurs confèrent une électronégativité et une capacité de complexation des cations. Dans le modèle de l'humification, les acides humiques dérivent des acides fulviques par polycondensation aléatoire. Ils possèdent un nucléus aromatique plus important et des chaînes latérales plus courtes. Les acides humiques forment des complexes avec les argiles, les minéraux et ions métalliques, ainsi qu'avec les composés organiques hydrophobes (Parent & Velegol, 2004). L'aboutissement de la polycondensation produit l'humine très stable et qui peut être solidement fixé sur les argiles ou gels colloïdaux.

Des composés lipidiques (acides gras, cires, résines) et azotés (ADN, chitine, muréine, urée, amines, protéines) sont également présents dans le sol. Ces composés hérités des bactéries, plantes ou animaux peuvent se fixer à des composés humifiés (Gobat *et al.*, 2003) ou à des colloïdes minéraux.

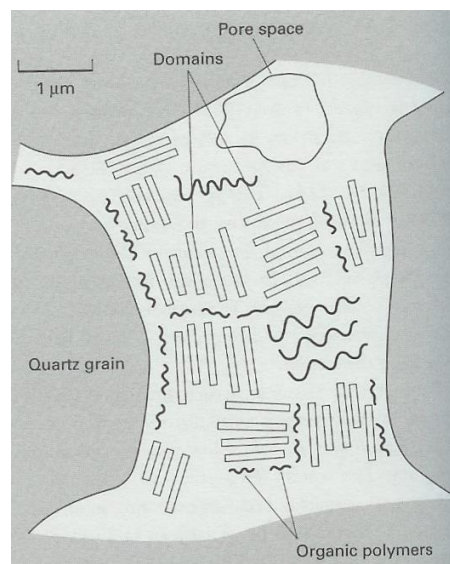
#### Structuration active par les organismes vivants

Le sol contient des êtres vivants appartenant à chacun des phylums du vivant : Bacteria, Archaeobacteria et Eucarya. Parmi les Eucaryotes, des algues, des champignons et des représentants de pratiquement tous les embranchements du monde animal sont retrouvés dans le sol. Chacun participe à la formation et à l'évolution du sol.

Les matières organiques et minérales sont brassées entre les zones superficielles et profondes par divers invertébrés et certains mammifères. La litière est fragmentée par les phytosaprophages. Les matières organiques alors dérivées présentent divers groupements fonctionnels, permettant leur liaison aux particules minérales. Leur interaction avec les particules minérales et la formation d'agrégats est favorisée lors du malaxage effectué dans le tube digestif des macroarthropodes et des vers de terre.



(A)



(B)

**Figure 4 : Les agrégats dans le sol** (White, 2006)

(A) Représentation schématique des macroagrégats dans le sol avec en (B) la représentation d'un arrangement possible des grains de quartz, des polymères organiques, et des argiles (domains) dans un microagrégat.

Les agrégats ainsi formés sont stabilisés par les réseaux formés par les hyphes des champignons, les sécrétions intestinales des invertébrés et des colloïdes bactériens qui servent de ciment (Gobat *et al.*, 2003).

#### Description des agrégats formés

Le sol est composé de structures agrégées de différentes formes, tailles et de stabilités variables. Ces structures sont formées suite aux propriétés physico-chimiques des argiles, mais également grâce à la présence de la matière organique, et l'activité de la faune (Gobat *et al.*, 2003). Les structures agrégées peuvent être divisées en deux grands types suivant leur taille (Millan *et al.*, 2006; White, 2006) (**Figure 4**)

Les **microagrégats** de 2 à 250  $\mu\text{m}$ , parmi lesquels sont distingués les microagrégats de 2 à 20  $\mu\text{m}$  très stables, formés de polysaccharides bactériens et de matières organiques fortement aromatiques liées aux argiles et aux limons fins. Des limons grossiers et des sables peuvent également être agrégés entre eux par des polysaccharides bactériens pour former des microagrégats plus gros, de 20 à 250  $\mu\text{m}$ .

Les **macroagrégats** de 50 à plus de 2 000  $\mu\text{m}$  sont formés de microagrégats reliés par des polysaccharides bactériens, des cellules bactériennes et du mycélium. Ces microagrégats peuvent être associés et consolidés entre eux par des particules de matière organique libre, des racines, et du mycélium (Gobat *et al.*, 2003). Le sol peut donc présenter de nombreux niveaux d'organisation, chaque niveau ayant une stabilité variable.

La stabilisation des agrégats dans le sol peut être corrélée au type de matière organique présente. L'association des polymères humiques avec les argiles est très stable, les polysaccharides bactériens participant à la formation des microagrégats et macroagrégats vont donner des structures de persistantes à transitoires. Les racines et les hyphes sont quant à eux responsables de macroagrégats transitoires (White, 2006).

### **1.3 Porosité, eau, et oxygénation**

La formation des agrégats est à l'origine de la création d'espaces vides qui augmentent la porosité du sol. La porosité est une composante de la modélisation des processus physico-chimiques et biologiques dans tous les types de sol (Schurgers *et al.*, 2006). Les pores ont une taille variable, allant de moins de 0,2  $\mu\text{m}$  jusqu'à plus de 50  $\mu\text{m}$  de diamètre (Gobat *et al.*, 2003). Dans les sols ayant une structure agrégée, la porosité peut être divisée en deux classes (Ranjard & Richaume, 2001; Rappoldt & Crawford, 1999) : Les microporosités inférieures à

10 µm, généralement situées dans les microagrégats ; et les macroporosités supérieures à 10 µm, localisées entre et dans les macroagrégats (Ranjard & Richaume, 2001).

La porosité du sol définit le volume du réservoir hydrique du sol (Gobat *et al.*, 2003). En condition non saturante, seule une fraction de la porosité est occupée par la solution du sol. Dans cette eau sont dissous de nombreux solutés issus de l'altération des roches et du métabolisme des êtres vivants. Le reste constitue l'atmosphère du sol. Deux paramètres importants de la porosité vont conditionner l'occupation des pores par l'eau ou par l'air : la distribution de la taille des pores et la topologie du réseau / espace poreux (Gimenez *et al.*, 1997; Rappoldt & Crawford, 1999; Vogel & Roth, 2001).

L'espace poreux doit en effet être considéré dans un espace tridimensionnel, avec des connections ou non entre pores de différentes tailles. La connectivité des pores et la tortuosité sont ainsi deux paramètres pouvant influencer les propriétés hydrauliques d'un sol, tels que la vitesse d'infiltration de l'eau et le transport de solutés dans le sol (Gimenez *et al.*, 1997; Vogel & Roth, 2001). La capacité de rétention de l'eau dans un sol, quant à elle, est surtout dépendante de la distribution des pores dans un sol (Gobat *et al.*, 2003; Schurgers *et al.*, 2006). Alors que l'eau est facilement drainée par gravité dans la macroporosité (surnommée porosité pour l'air), elle est plus ou moins fortement retenue dans les pores de moins de 50µm par capillarité (Gobat *et al.*, 2003). Ainsi les sols de type sableux ont une plus faible capacité de rétention que les sols argileux (Schurgers *et al.*, 2006).

La rétention de l'eau est liée d'une part aux effets de liaison de l'eau avec la matrice, d'autre part à l'effet capillaire dans les micropores qui présentent des tensions de surface élevées (Gobat *et al.*, 2003; Schurgers *et al.*, 2006). La combinaison de ces deux forces dans la microporosité étant supérieure à la force de gravité, l'eau est retenue. Cette solution du sol peut être utilisée par les organismes vivants (dont les bactéries) à condition de développer une force de succion supérieure à celle de la force de rétention (Gobat *et al.*, 2003). En contrepartie, les micropores sont moins sujets à des alternances saturation / dessiccation que les macroporosités (Ranjard & Richaume, 2001).

Les conditions d'oxygénation de l'espace poral dépendent de la connectivité des pores vides avec l'atmosphère, de la solubilité ainsi que de la diffusion de l'oxygène dans l'eau (Rappoldt & Crawford, 1999), mais également de sa consommation par respiration (Rappoldt & Crawford, 1999; Schurgers *et al.*, 2006) et de la capacité de rétention d'eau dans le sol (Schurgers *et al.*, 2006). Du fait de la capacité de rétention d'eau du sol, un certain nombre de pores sont occupés par la solution du sol.

La diffusion de l'oxygène dans l'eau est 10 000 fois inférieure à celle dans l'air, sans compter sa faible solubilité (Rappoldt & Crawford, 1999). L'absence de connectivité entre pores aérés ou avec l'atmosphère pose la question des conditions aérobies ou non dans ces pores. Selon certains modèles, des conditions anoxiques devraient rapidement apparaître dans les pores « vides » non connectés à cause de la respiration, de la faible diffusion de l'oxygène dans l'eau, et de la tortuosité du réseau poreux (Rappoldt & Crawford, 1999). Schurger *et al.* (2006) considèrent que l'anoxie suscite et peut induire des processus de respiration anaérobie (dénitrification par exemple) lors d'une augmentation de l'humidité du sol. Les autres gaz, comme le dioxyde de carbone, les oxydes d'azotes ou des gaz toxiques pour les microorganismes peuvent entrer dans la composition de l'atmosphère du sol. Ils sont également soumis aux processus de diffusion et aux mêmes lois physiques évoquées pour l'oxygène.

#### **1.4 Présence de microenvironnements dans le sol**

Les conditions physico-chimiques et biologiques résultant de l'aggrégation, de l'hétérogénéité spatiale de l'aération et de la biodisponibilité des solutés conduisent à l'existence de microenvironnements.

Les micropores, peu sujets à des variations d'humidité grâce aux forces de rétention, forment une niche protectrice contre la dessiccation. Leur saturation en eau entraîne des conditions proches de l'anoxie. Ces sites contiennent une forte concentration en matière organique, dérivée des matières humifiées difficilement dégradables. La présence des argiles ayant une forte capacité échangeuse d'ions est importante pour les bactéries : elle permet le stockage des nutriments grâce à sa capacité d'échange, tout en les laissant biodisponibles pour les bactéries (Young & Crawford, 2004).

Les macropores sont plus sujets à dessiccation, mais possèdent également une meilleure conductivité hydraulique et donc un transport des solutés plus important et plus rapide. Les conditions y sont le plus souvent oxiques, sauf lors des épisodes de saturation (sous-sol imperméable, fortes pluies, irrigation).

## **2 LES BACTERIES DANS L'ENVIRONNEMENT SOL**

Des dénombrements effectués sous microscope permettent d'estimer la densité bactérienne à  $10^8$ - $10^{10}$  cellules par gramme de sol sec (Alexander, 1977). Ces bactéries sont rarement présentes à l'état libre dans la phase liquide du sol (Alexander, 1977). Plus de 80-90 % de la fraction bactérienne serait attachée (Bakken, 1985; Nannipieri *et al.*, 2003; Steffan *et al.*, 1988). L'attachement ou la formation de biofilm prolongerait la survie des cellules dans le sol (Chen & Zhu, 2005). Elle aiderait les cellules à concentrer les nutriments via l'effet interfacial sur les particules minérales (Walker *et al.*, 1989). Cet attachement n'est pas uniforme dans la structure sol, ni irréversible. Des méthodes adaptées ont permis d'en apprendre un peu plus sur la distribution des bactéries dans l'environnement sol, ainsi que sur les mécanismes intervenant dans leur attachement aux particules minérales.

### **2.1 Distribution des bactéries dans le sol**

Pour analyser les agrégats en tant qu'habitat potentiel des bactéries, une méthode consiste tout d'abord à disperser les différents agrégats dans de l'eau. Les agrégats les plus stables sont ensuite séparés suivant leur taille par tamisage. Des différences de population bactérienne ont été observées entre la partie interne des microagrégats, la totalité des microagrégats, et les macroagrégats (Mummey *et al.*, 2006; Mummey & Stahl, 2004). Des différences de populations fixatrices d'azote, plus ou moins marquées suivant les sols analysés, ont également été observées entre différentes tailles d'agrégats (de la fraction argileuse au macroagrégat) (Izquierdo & Nusslein, 2006). Une analyse de la littérature effectuée par Ranjard *et al.* (2001) indique que la densité bactérienne la plus haute a été observée dans les agrégats inférieurs à 50  $\mu\text{m}$ . En effet, 80 % des bactéries seraient localisées dans la partie interne des agrégats. Les bactéries sont soit localisés dans les pores, soit enlisés dans l'argile (Foster, 1988; Grundmann, 2004), avec une colonisation préférentielle dans les pores de 2  $\mu\text{m}$  (Ranjard & Richaume, 2001).

Une analyse plus fine à micro-échelle a montré des augmentations locales de la densité bactérienne (Dechesne *et al.*, 2003; Gonod *et al.*, 2003; Grundmann *et al.*, 2001; Nunan *et al.*, 2002; Nunan *et al.*, 2003). Ces agrégats bactériens sont plus ou moins denses avec parfois seulement 2 à 5 cellules bactériennes sur une particule minérale (Grundmann, 2004). Une observation microscopique de certains de ces agrégats bactériens a montré une grande diversité morphologique dans ces microsites (Nunan *et al.*, 2003). Ces regroupements bactériens peuvent être expliqués par un microenvironnement particulier favorisant le

développement des cellules (source locale de nutriments, pH modifié localement, production / sécrétion de molécules stimulatrices, consortium bactérien) (Nunan *et al.*, 2003).

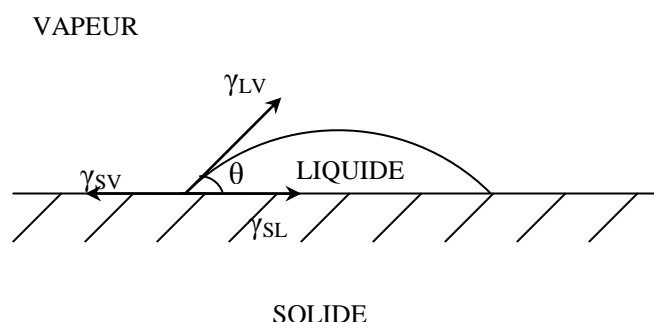
Certaines études ont ainsi montré l'influence du nutriment sur la distribution bactérienne. Une étude réalisée sur des sols labourés a ainsi montré une distribution aléatoire d'agrégats bactériens (très denses) dans la zone superficielle (labourée), alors que ceux (moins denses) des zones plus profondes sont localisés dans le voisinage immédiat des pores (Nunan *et al.*, 2003). Cette modification de distribution serait liée à la localisation des substrats nutritifs (matières organiques). Les substrats organiques seraient uniformément répartis dans la zone superficielle du sol labouré. La fraction soluble serait mobilisée lors d'infiltration d'eau. Ceci expliquerait une distribution préférentielle des bactéries à proximité des pores dans les zones plus profondes (Nunan *et al.*, 2003). De même, une augmentation de la densité bactérienne et une localisation préférentielle à proximité des excréments de la mésofaune ont été constatées (Bruneau *et al.*, 2005).

Les stratégies de distribution varient suivant les espèces bactériennes. Par exemple, il a été observé que les populations dégradant l'acide 2,4 dichlorophénoxyacétique (2,4D) forment en effet des agrégats clairement distincts, peu denses et très espacés, alors que les populations oxydant le  $\text{NH}_4^+$  forment dans la même expérience des petits agrégats plus denses et moins espacés (Dechesne *et al.*, 2003). Cette variation de distribution peut être liée aux stratégies « adaptatives » ou nutritives des populations considérées. En analysant des unités de volume de sol de tailles différentes (20 à 500  $\mu\text{m}$ ), Grundmann *et al.* (2001) ont noté un profil spatial encore plus diffus des populations oxydant le nitrite ( $\text{NO}_2^-$ ) par rapport aux populations oxydant le  $\text{NH}_4^+$ . Ces deux populations nitrifiantes restent toutefois spatialement dépendantes. Cette distribution plus diffuse mais non aléatoire permettrait une meilleure interception des ions  $\text{NO}_2^-$  par les populations oxydant le  $\text{NO}_2^-$  (Grundmann *et al.*, 2001).

## 2.2 Les propriétés de la surface bactérienne

La cellule bactérienne est délimitée de l'environnement extérieur par une enveloppe composée de la membrane cytoplasmique, d'une couche de peptidoglycane plus ou moins épaisse, surmontée (chez les protéobactéries) ou non (cas des bactéries Gram positifs) d'une deuxième membrane appelée membrane externe. De la composition de la surface externe des cellules bactériennes vont dépendre les propriétés de surface des bactéries (Strevett & Chen, 2003).





Espèce	Angle de contact $\theta$ (°)
<i>Pseudomonas fluorescens</i>	21,2 +/- 1,5
<i>Pseudomonas aeruginosa</i>	25,7 +/- 0,9
<i>Pseudomonas putida</i>	38,5 +/- 1,0
<i>Pseudomonas</i> sp. strain 26-3	20,1 +/- 0,8
<i>Pseudomonas</i> sp. strain 52	19,0 +/- 1,0
<i>Pseudomonas</i> sp. strain 80	29,5 +/- 0,5
<i>Escherichia coli</i> NCTC 9002	15,7 +/- 1,2
<i>Escherichia coli</i> K-12	24,7 +/- 0,4
<i>Arthrobacter globiformis</i>	23,1 +/- 0,7
<i>Arthrobacter simplex</i>	37,0 +/- 0,9
<i>Arthrobacter</i> sp. strain 177	60,0 +/- 1,5
<i>Arthrobacter</i> sp. strain 127	38,0 +/- 1,3
<i>Micrococcus luteus</i>	44,7 +/- 0,9
<i>Acinetobacter</i> sp. strain 210A	32,6 +/- 0,5
<i>Thiobacillus versutus</i>	26,8 +/- 0,8
<i>Alcaligenes</i> sp. strain 175	24,4 +/- 0,5
<i>Rhodopseudomonas palustris</i>	34,3 +/- 0,5
<i>Agrobacterium radiobacter</i>	44,1 +/- 0,5
<i>Bacillus licheniformis</i>	32,6 +/- 0,5
<i>Corynebacter</i> sp. strain 125	70,0 +/- 3,0
<i>Azotobacter vinelandii</i>	43,8 +/- 0,5
<i>Rhizobium leguminosarum</i>	31,0 +/- 1,0
<i>Mycobacter phlei</i>	70,0 +/- 5,0

**Figure 5 : Hydrophobicité de différentes espèces bactériennes** (Van Loosdrecht, 1987)

De la même manière que l'hydrophobicité d'un solide est mesurée par l'angle de contact  $\theta$  entre l'eau et le solide, l'hydrophobicité de la surface bactérienne est mesurée par l'angle de contact de l'eau sur la cellule bactérienne. L'augmentation de  $\theta$  correspond à une augmentation de l'hydrophobicité de la surface de la cellule bactérienne. En dessous de  $30^\circ$ , les bactéries n'adhèrent pas à une phase apolaire. Au dessus de  $30^\circ$ , l'augmentation de l'angle de contact coïncide avec une augmentation de l'adhésion des cellules bactériennes.

Celles-ci possèdent un caractère hydrophobe, plus ou moins prononcé suivant les espèces bactériennes (Doyle, 2000; Rosenberg, 2006; Stenstrom, 1989; Strevett & Chen, 2003; van Loosdrecht *et al.*, 1987) (**Figure 5**). Il peut être dû à la présence de protéines hydrophobes (hydrophobines) associées à la paroi, liées à la membrane externe ou à la membrane cytoplasmique (Doyle, 2000). L'hydrophobicité est également liée aux résidus apolaires des fimbriae / pilis (Donlan, 2002; Doyle, 2000) et des exopolysaccharides (EPS) sécrétés par les bactéries (Doyle, 2000). D'autres composantes trouvées spécifiquement chez certaines bactéries comme les acides lipotéiques et mycoliques produits par les bactéries à Gram positif, ou la capsule des spores chez les bactéries capables de sporuler (Doyle, 2000) participent également au caractère hydrophobe des surfaces bactériennes. Au niveau moléculaire, le caractère hydrophobe est dû à la présence de longues chaînes carbonées apolaires. Cette hydrophobicité augmente avec le nombre de doubles ou triples liaisons entre deux carbones et la délocalisation possible des électrons.

En plus des résidus hydrophobes, les surfaces bactériennes comportent des résidus chargés (Doyle, 2000). L'analyse de la surface bactérienne par spectrométrie Infra Rouge (IR) révèle en effet une grande variété de groupes fonctionnels tels que des groupements éthers, cétones, aldéhydes, hydroxyles, carbonyles et carboxyles qui génèrent une électronégativité générale de la cellule bactérienne et lui apportent de ce fait un caractère hydrophile. Cette hydrophilie est due à l'antigène O des lipopolysaccharides (LPS), et aux EPS sécrétés par les bactéries (Donlan, 2002). Outre les charges négatives, les bactéries présentent également des charges positives localisées (Stenstrom, 1989).

Grâce à l'extension du modèle Derjaguin-Landau-Verwey-Overbeck ou DVLO (utilisé pour décrire les tensions de surface des colloïdes), certains auteurs sont capables d'expliquer le comportement des surfaces bactériennes. En tenant compte des différentes interactions potentielles, Strevett & Chen (2003) déterminent le caractère globalement hydrophile ou hydrophobe de la surface bactérienne. Deux types d'interactions sont prises en compte : les liaisons Van der Waals entre résidus hydrophobes, et les interactions entre acides et bases de Lewis à l'origine de liaison covalente. Ce caractère global est variable d'une espèce à l'autre, voire même au sein d'une espèce. Il peut varier selon les conditions de croissance de la bactérie (Stenstrom, 1989; Strevett & Chen, 2003), l'état physiologique des cellules (Strevett & Chen, 2003) et les structures microbiennes de surface produites par la bactérie (Strevett & Chen, 2003). Ainsi, les cellules bactériennes possédant une surface bactérienne globalement hydrophile auront tendance à être dispersées en milieux aqueux, alors que celles présentant une surface hydrophobe auront tendance à s'agréger. En pratique,

les cellules bactériennes sont diluées dans la matrice sol, la probabilité de leur rencontre est donc faible. La probabilité de collision avec la surface des particules sableuses est par contre 1 000 à 10 000 fois supérieure à celle des cellules, conduisant à un attachement à la surface minérale préférentiellement à une agrégation entre cellules (Chen & Zhu, 2005).

En plus du caractère global (hydrophobe ou hydrophile) des surfaces bactériennes qui déterminent l'attachement sur des surfaces telles que les particules de sables, les interactions électrostatiques peuvent devenir un paramètre important lors de l'attachement de la cellule bactérienne à une surface chargée (argile, humine). Elles tiennent compte du potentiel global des deux surfaces qui entrent en contact. Les cellules bactériennes et les particules de sol étant toutes deux chargées négativement, une répulsion peut avoir lieu entre les deux surfaces. L'attachement ou non à la surface dépendra ainsi de la force des deux autres interactions (Van der Waals, acide / base de Lewis) qui auront tendance à attirer les deux surfaces. Suivant la prédominance des forces, l'attachement aura lieu ou non (Strevett & Chen, 2003). Ainsi, le processus d'adhésion se déroule en deux étapes : Au cours de la première approche, les forces d'interaction doivent être supérieures à la diffusion ou dispersion hydrodynamique ce qui entraîne une adsorption initiale. Suivant la force des interactions, la répulsion électrostatique est contre-balancée ou non, ce qui entraîne soit une remise en suspension, soit un rapprochement de la surface. A ce stade, l'adhésion est réversible. Dans une deuxième phase d'approche, les interactions électrostatiques diminuent jusqu'à devenir négligeables alors que les interactions acide / base de Lewis et Van der Waals prennent de l'ampleur jusqu'à la liaison et l'attachement irréversible à la surface (Chen & Zhu, 2005; Liu *et al.*, 2004; Strevett & Chen, 2003).

Des expérimentations d'attachement des bactéries à des particules de la fraction sableuse ont été réalisées (Stenstrom, 1989). Ces essais ont montré un parallèle entre l'augmentation de l'hydrophobicité des cellules et l'adhésion aux particules minérales de quartz et feldspath. L'effet répulsif lié aux charges négatives globales des surfaces bactériennes et des particules minérales est beaucoup moins clair. Quelque soit le potentiel négatif des particules de sol, aucune modification de l'adhésion des bactéries n'a été observée. Les charges positives présentes localement sur les particules de sol montrent par contre un effet beaucoup plus marqué que les charges négatives. Les auteurs ont en effet observé une corrélation entre la présence de charges positives et l'adhésion aux particules minérales. L'hydrophobicité reste cependant le principal « moteur » dans l'attachement des bactéries (Stenstrom, 1989). Les résultats obtenus sont en accord avec les modèles thermodynamiques et d'autres études effectuées sur des fractions sableuses (Chen & Zhu, 2005; Liu *et al.*, 2004).

Lorsque l'on s'intéresse aux argiles, l'attachement des cellules bactériennes semble plus conditionné par les interactions électrostatiques. Une étude de l'attachement de *Pseudomonas putida* sur différentes argiles a montré une adsorption plus élevée sur les minéraux chargés positivement (Goethite) et un attachement le plus faible pour les argiles comportant le plus de charges négatives (montmorillonite versus kaolinite) (Jiang *et al.*, 2007). Par ailleurs, une diminution de l'adsorption est notée pour une augmentation concomitante de la teneur du sodium et du pH. L'adhésion de *Pseudomonas putida* sur les argiles est diminuée de plus de 50 % lors d'une augmentation du pH (argiles chargées négativement au dessus de leur point zéro charge PZC). Contrairement à l'étude de Stenstrom *et al.* (1989) sur les sables, l'adsorption des cellules sur les argiles est fortement influencé par le pH et par les cations de la capacité d'échange cationique.

### **2.3 Accès aux cellules bactériennes telluriques**

Les bactéries sont donc distribuées de manière hétérogène dans la matrice sol, selon les conditions environnementales et les stratégies « nutritives » propres à chaque espèce bactérienne. La majorité des cellules bactériennes étant attachées aux particules de sol (Bakken, 1985; Nannipieri *et al.*, 2003; Steffan *et al.*, 1988) avec une distribution préférentielle dans les microporosités (Ranjard & Richaume, 2001), une déstructuration du sol et une désorption des cellules de la matrice sol sont nécessaires pour recouvrer les diverses bactéries présentes dans l'échantillon étudié. Sans déstructuration préalable de la matrice sol, 67 à 92 % des cellules resteraient inaccessibles (Frostegard *et al.*, 1999; Gabor *et al.*, 2003; Zhou *et al.*, 1996).

Une méthode de déstructuration consiste à disperser le sol dans un tampon phosphate par broyage au Waring Blender, méthode de disruption ayant permis le meilleur recouvrement des cellules bactériennes comparativement à la sonication ou à des traitements chimiques (Lindahl, 1996). Le type de sol a cependant un effet important sur le recouvrement des cellules (Gabor *et al.*, 2003). Pour des sols sableux, 79 % des cellules sont extraites, 55 % pour des sols limoneux et moins de 29 % pour des sédiments / boues marines. La fixation des bactéries sur les particules de sol semble avoir un rôle non négligeable. Les rendements d'extraction des cellules bactériennes sont en effet d'autant plus faibles que la quantité d'argile augmente dans un sol (Bakken, 1985).

Le tampon pyrophosphate aide à la dispersion des cellules du sol. L'utilisation de ce tampon augmente le nombre de cellules détectées en solution. Il permettrait d'une part la disruption des agrégats par le relâchement des acides humiques. Ce tampon agit d'autre part sur les propriétés d'agrégation des bactéries. Aucune agrégation n'est en effet observée dans ce tampon contrairement à une dispersion dans l'eau. Le tampon altérerait la matrice exopolysaccharidique (EPS) des bactéries (Lindahl, 1996). Des enzymes dégradant les exopolysaccharides peuvent également être utilisées pour favoriser le détachement des cellules des particules sableuses. Bockelmann *et al.* (2003) ont effectivement observé une amélioration de l'extraction des cellules du sol via un traitement enzymatique par une  $\alpha$ -glucosidase et une  $\beta$ -galactosidase, ainsi qu'une lipase pour différents sols d'Allemagne (Bockelmann *et al.*, 2003).

D'autres méthodes disruptives de la matrice sol peuvent être utilisées, mais sont cependant beaucoup plus destructrices que le Waring Blender. Le traitement Waring Blender peut léser l'intégrité cellulaire lors d'une utilisation prolongée (Lindahl, 1996). C'est pourquoi la plupart des protocoles préconisent des temps courts répétés (3 x 1 min) pour la dispersion du sol dans le tampon phosphate, avant extraction des cellules de la matrice sol. La méthode disruptive par broyage à l'aide de billes (silice, acier) permet une bonne déstructuration du sol mais entraîne également une lyse des bactéries. L'ADN cellulaire est alors libéré dans la matrice sol et peut interagir avec les surfaces réactives. L'accès à la microflore tellurique via cette technique implique donc la lyse *in situ* des bactéries dans des conditions limitant l'adhésion aux surfaces (tampon dispersant) suivie immédiatement de l'extraction de l'ADN de la matrice sol.

### **3 LES MOLECULES ADN DANS LA MATRICE SOL**

La première voie pour accéder à la microflore tellurique passe la lyse des bactéries conjointement à la déstructuration du sol, via des traitements physiques et chimiques drastiques. Une telle approche requiert l'extraction subséquente de l'ADN libéré de la matrice sol et sa purification. Toutefois, la validation d'une telle stratégie nécessite la prise en compte de l'ADN extracellulaire déjà présent dans le sol et des interactions entre ADN libéré par les traitements lytiques et les composants du sol.

### 3.1 L'ADN extracellulaire dans le sol

De l'ADN à l'état libre (extracellulaire) a été détecté dans le sol à raison de 4 à 17  $\mu\text{g.g}^{-1}$  de sol (Courtois *et al.*, 2001; Frostegard *et al.*, 1999). Cet ADN extracellulaire est en partie originaire d'organismes morts présents dans le sol, que ceux-ci appartiennent aux règnes des Eucaryotes (plantes animaux et champignons) des Procaryotes ou des Archaea (Gobat *et al.*, 2003). Chez certaines bactéries dont *Acinetobacter baylyi* (anciennement nommée *A. calcoaceticus*), la phase exponentielle de croissance s'accompagne d'un taux élevé de mort cellulaire et donc de libération d'ADN plasmidique et chromosomique (Lorenz *et al.*, 1991). Cet ADN devient disponible pour le reste de la population qui développe un stade de compétence et peut internaliser l'ADN extracellulaire.

Des cellules intactes vivantes peuvent également excréter de l'ADN. Ce phénomène a été observé chez de nombreux taxons bactériens (Draghi & Turner, 2006; Lorenz & Wackernagel, 1994; Paget & Simonet, 1994). Une excrétion présumée active d'ADN chromosomique et plasmidique a ainsi été observée chez *Bacillus subtilis* pendant sa phase de compétence (Lorenz *et al.*, 1991). Dans le cas de *Neisseria meningitidis*, souche libérant également de l'ADN, un îlot génomique conférant une sécrétion active d'ADN par sécrétion de type IV a été détectée dans 80 % des souches séquencées. Le coût énergétique ainsi investi dans l'excrétion d'ADN a été interprété comme apportant un bénéfice pour la population via un brassage génétique par échanges d'allèles (Draghi & Turner, 2006; Lorenz *et al.*, 1991). Certaines souches, comme *Pseudomonas stutzeri*, ont en effet une transformation augmentée au contact de cet ADN par rapport à de l'ADN purifié ou de l'ADN provenant de la lyse cellulaire (Stewart *et al.*, 1983).

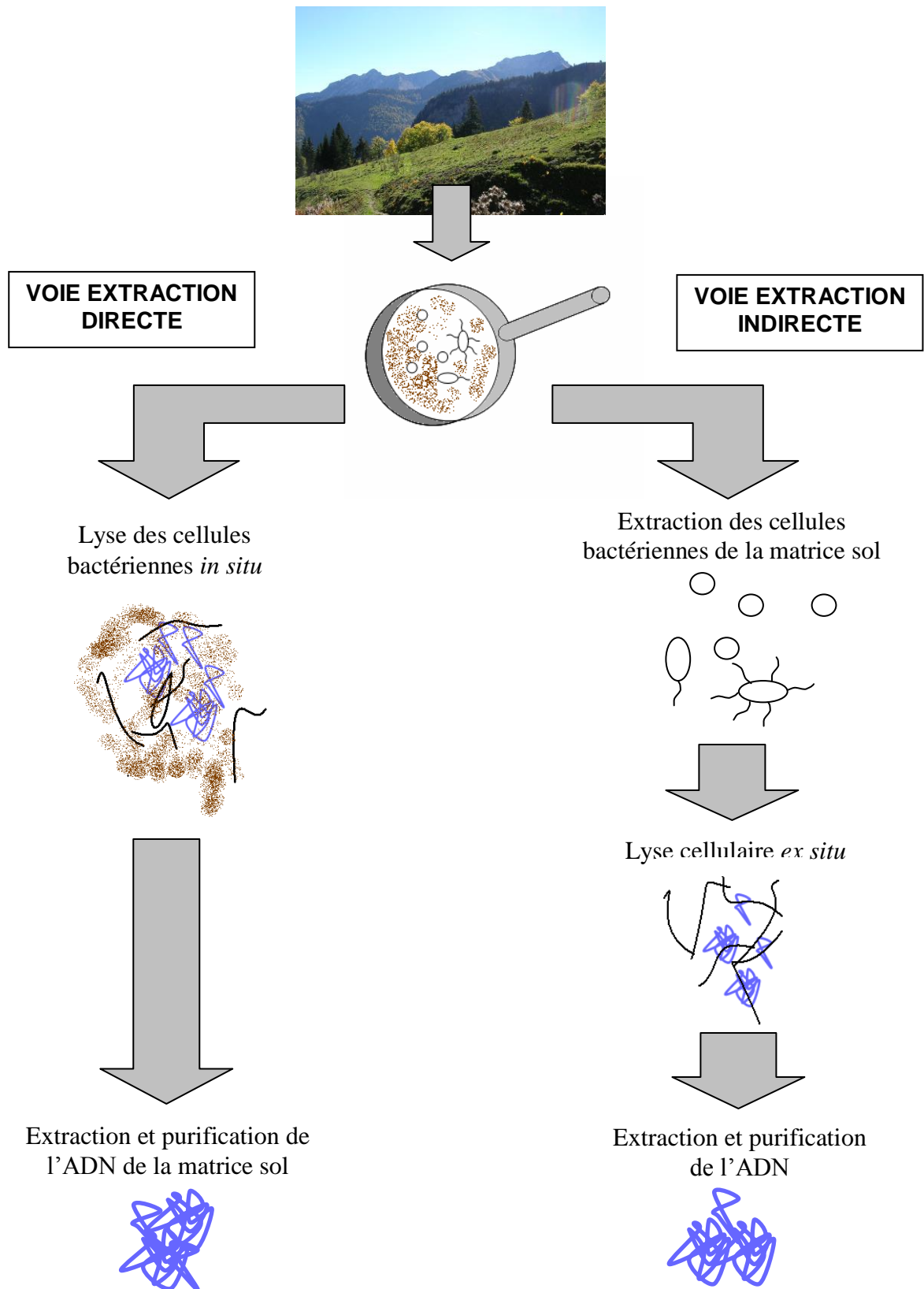
Cet ADN extracellulaire peut persister dans le sol, via un effet protecteur apporté par les débris cellulaires (Ceremonie *et al.*, 2004; Nielsen *et al.*, 2000), ou grâce à son adsorption sur les particules de sol (Romanowski *et al.*, 1992). Plusieurs études montrent en effet une protection de l'ADN fixé sur les particules de sol contre les activités nucléasiques (Demaneche *et al.*, 2001b; Romanowski *et al.*, 1992). Cet effet protecteur des particules de sol serait en partie expliqué par l'adsorption des enzymes sur la matrice sol et sa séparation physique de l'ADN (Demaneche *et al.*, 2001b).

### 3.2 Interaction entre ADN et composants du sol

Dans un microcosme simulant des conditions environnementales, 80 % des acides nucléiques introduits sont adsorbés sur des sédiments aquifères en moins de 20 min (Romanowski *et al.*, 1993). Des protocoles permettent l'extraction de l'ADN de la matrice sol, les rendements d'extraction sont cependant beaucoup plus faibles pour les sols riches en argiles (Frostegard *et al.*, 1999; Zhou *et al.*, 1996). L'ADN est en effet une molécule polyanionique. Le groupement phosphate des bases azotées confère à la molécule d'ADN une charge négative, avec une réactivité variable suivant les conditions pH existant dans le sol. Suite aux interactions électrostatiques, les molécules d'ADN s'adsorbent sur les particules de sol, notamment sur les argiles chargées positivement.

L'adsorption de l'ADN est influencée par la CEC (capacité d'échange cationique) des particules (Jiang *et al.*, 2007). Cette adsorption est augmentée en présence d'ions divalents ( $\text{Mg}^{2+}$ ,  $\text{Ca}^{2+}$ ) voire multivalents, par rapport aux ions monovalents ( $\text{Na}^+$ ,  $\text{K}^+$ ) (Chamier *et al.*, 1993; Lorenz & Wackernagel, 1994). Cette adsorption est particulièrement efficace avec les ions  $\text{Ca}^{2+}$  comparativement aux ions  $\text{Mg}^{2+}$  (Chamier *et al.*, 1993; Paget & Simonet, 1994). La plus forte solubilité des ions  $\text{Mg}^{2+}$  sous forme de phosphate de magnésium pourrait expliquer ce phénomène (Paget & Simonet, 1994). Ces cations serviraient de médiateurs pour la complexation entre les groupements phosphates de l'ADN et les charges négatives des minéraux.

L'extraction de l'ADN libéré dans le sol, alors adsorbé sur les particules d'argiles, nécessite la mise en oeuvre de traitements permettant la désorption de l'ADN. Des travaux conduits sur des systèmes simplifiés, composés de particules d'argiles pures et d'ADN, ont montré le rôle de sels sur la désorption des molécules d'ADN (Demaneche *et al.*, 2001b). Par exemple, lorsqu'un complexe ADN-argile est placé dans un milieu riche en NaCl tel que le milieu LB (NaCl à  $10 \text{ g.L}^{-1}$ ), de l'ADN est détecté en solution. Pour un milieu contenant moitié moins de sel (LBm) ou simplement dans de l'eau, aucune désorption d'ADN n'est constatée (Demaneche *et al.*, 2001b). Cette désorption serait liée au remplacement sur les surfaces minérales des ions  $\text{Ca}^{2+}$  et  $\text{Mg}^{2+}$  par l'ion  $\text{Na}^+$ . Ainsi les tampons d'extraction riches en sels favorisent la désorption de l'ADN des particules de sol (Ranjard *et al.*, 1998).



**Figure 6 :** Représentation schématique des deux voies d'extraction de l'ADN des bactéries contenues dans un échantillon de sol



L'ajout d'ARN dans les tampons d'extraction permet également l'augmentation de la quantité d'ADN extraite. L'ARN est un compétiteur de l'ADN pour les sites d'adsorption sur les particules d'argiles, expliquant ainsi l'augmentation des rendements postérieurement à l'introduction d'ARN dans le sol (Frostegard *et al.*, 1999). L'ADN présent dans la matrice sol est ainsi désorbé quelque soit son origine.

## **4 EXTRACTION DU METAGENOME**

Les approches métagénomiques sont basées sur l'exploitation du patrimoine génétique bactérien extrait directement des cellules vivant dans leur milieu naturel, indépendamment de toute étape de mise en culture. Ces approches ont nécessité le développement de méthodes recherchant la meilleure efficacité en terme d'exhaustivité et de représentativité de l'ADN extrait, d'intégrité physique des molécules d'ADN et de pureté des solutions d'extractions. L'environnement sol implique la prise en compte de la distribution hétérogène des microorganismes dans la structure du sol, des interactions des cellules bactériennes ou de l'ADN avec les composants du sol, ainsi que de tous les contaminants pouvant être conjointement extraits avec le métagénome bactérien. Deux voies d'extraction du métagénome sont utilisées suivant la lyse cellulaire *in situ* (voie d'extraction directe) ou une lyse *ex situ* postérieurement à l'extraction cellulaire (voie d'extraction indirecte) (**Figure 6**). Les méthodes d'extractions, décrites en détail dans l'article de Robe *et al.* (2003), sont présentées ci-dessous et comparées.

### **4.1 Isolement de l'ADN après lyse cellulaire *in situ***

Dans cette approche, les cellules sont généralement lysées conjointement à la déstructuration du sol. Les approches faisant intervenir la disruption physique par agitation de billes d'acier donnent notamment de meilleurs rendements que les chocs thermiques par congélation / décongélation, l'action des micro-ondes, ou le broyage d'échantillons dans l'azote liquide (Cullen & Hirsch, 1998; Kauffmann *et al.*, 2004; Lakay *et al.*, 2007; Leff *et al.*, 1995). L'action vigoureuse des billes permet notamment la disruption cellulaire de bactéries à Gram positif, plus réfractaire à la lyse que les protéobactéries (Frostegard *et al.*, 1999; Kauffmann *et al.*, 2004). C'est toutefois par la combinaison de procédés physique, chimique et enzymatique que de meilleures efficacités lytiques ont été atteintes. Ainsi des traitements enzymatiques via le lysozyme (Frostegard *et al.*, 1999; Ranjard *et al.*, 1998; Zhou *et al.*, 1996) pour son action hydrolytique sur les liaisons glycosidiques (Robe *et al.*, 2003), de

la protéinase K (Zhou *et al.*, 1996) et de l'achromopeptidase efficace sur les actinobactéries réfractaires (Courtois *et al.*, 2001; Simonet *et al.*, 1984) sont effectués pour favoriser la lyse cellulaire de l'ensemble de la communauté microbienne. Plus de 90 % des cellules bactériennes seraient ainsi lysées *in situ* (Howeler *et al.*, 2003; Ogram *et al.*, 1987; Zhou *et al.*, 1996).

L'ADN libéré est alors adsorbé sur les particules de sol, et plus particulièrement sur les argiles (Frostegard *et al.*, 1999; Zhou *et al.*, 1996). Les traitements d'homogénéisation contribueraient par la mise en contact de l'ADN avec les argiles à accroître le niveau d'adsorption et ainsi diminuer les rendements d'extraction (Frostegard *et al.*, 1999). Les tampons d'extraction doivent donc permettre la désorption des molécules d'ADN de la matrice sol. Ces difficultés peuvent être plus ou moins limitées par l'emploi de tampons riches en sels (Ranjard *et al.*, 1998) ou par la saturation des sites d'adsorption via l'ajout d'ARN (Frostegard *et al.*, 1999). Cependant, toutes les molécules d'ADN initialement présentes avant lyse *in situ* sont également désorbées. De l'ADN eucaryotique a ainsi été détecté dans les extraits d'ADN (Courtois *et al.*, 2001; Ranjard *et al.*, 1998); et constituerait plus de 60 % de l'ADN extrait par cette voie d'extraction (Gabor *et al.*, 2003). Par ailleurs, les acides humiques présentant des caractéristiques physico-chimiques similaires à l'ADN (macromolécules chargées négativement) sont également co-extraits.

Ces acides humiques interfèrent avec de nombreuses réactions enzymatiques (Wilson, 1997) et affectent ainsi les étapes ultérieures de manipulation de l'ADN. Des étapes de précipitation / purification (éthanolique, PEG<sup>a</sup>, CTAB<sup>b</sup>, PVPP<sup>c</sup>) ne permettent que leur élimination partielle. Les sols riches en acides humiques nécessitent une purification supplémentaire par séparation des macromolécules sur gradient de chlorure de césium où les molécules sont séparées selon leur densité, sur colonne d'hydroxyapatite, ou colonne de Séphadex ou colonne de Sépharose (Cullen & Hirsch, 1998; Lakay *et al.*, 2007; Ogram *et al.*, 1987). D'autres solutions ont été apportées par lavages préalables des sols avant extraction (Fortin *et al.*, 2004; Lakay *et al.*, 2007). L'utilisation avant lyse *in situ* par exemple de sulfate d'aluminium permet notamment la précipitation des acides humiques et ADN extracellulaires présents dans la matrice sol (Dong *et al.*, 2006).

---

<sup>a</sup> Polyéthylène glycol

<sup>b</sup> Cetyltriméthylammoniumbromide

<sup>c</sup> Polyvinylpyrrolidone

En conclusion, la recherche de la meilleure efficacité en terme de quantité d'ADN extrait selon la voie directe nécessite l'emploi de traitements lytiques drastiques qui en contrepartie vont avoir un effet délétère sur l'intégrité de l'ADN. Par exemple, l'emploi d'un traitement physique par agitation de billes d'acier va produire une solution d'ADN dans laquelle les fragments d'ADN ne dépassent pas 20 kb en taille, avec parfois des fragments de seulement 0,5 kb (Cullen & Hirsch, 1998; Frostegard *et al.*, 1999; Ogram *et al.*, 1987). Des valeurs publiées indiquent que rarement plus de 60 % de l'ADN bactérien est effectivement extrait du sol (Howeler *et al.*, 2003; Luna *et al.*, 2006). Près de 30 % de l'ADN est perdu lors des étapes de purification pour l'obtention d'un ADN de qualité suffisante en vue de son exploitation ultérieure (Howeler *et al.*, 2003). Il revient donc à l'expérimentateur de définir le traitement le plus approprié selon les caractéristiques physico-chimiques du sol, mais également des objectifs de l'expérimentation, pour un compromis entre quantité et qualité de l'ADN à extraire du sol.

#### **4.2 Isolement de l'ADN après lyse cellulaire *ex situ***

Les objectifs qui ont prévalu au développement de cette technologie visaient à s'affranchir des biais de la voie d'extraction directe liés à l'adsorption de l'ADN sur les composants du sol, aux difficultés rencontrées pour sa purification, et enfin à l'effet délétère des traitements sur l'état de l'ADN. Extraire préalablement les cellules bactériennes hors de la matrice devait permettre de purifier la suspension bactérienne de la plupart des composés du sol et en particulier des acides humiques. La suspension pouvait alors être soumise à des traitements lytiques à la fois efficaces pour lyser les cellules et ménagés pour ne pas dégrader l'ADN. En d'autres termes, la suspension bactérienne résultant de l'extraction à partir du sol pouvait être traitée selon des protocoles définis pour des cellules issues de culture pure donc parfaitement adaptés aux travaux ultérieurs de biologie moléculaire.

Pour l'extraction des cellules, le sol est préalablement désagréé. La méthode disruptive au Waring Blender dans un tampon phosphate est généralement utilisée, car elle permet un bon recouvrement des cellules bactériennes tout en conservant l'intégrité cellulaire (Lindahl & Bakken, 1995; Lindahl, 1996). Les cellules sont ensuite séparées des particules grossières par centrifugation différentielle, puis des particules plus fines et autres contaminants organiques (acides humiques) par un gradient de densité (Bakken, 1985). C'est avec le polymère Nycodenz que les gradients de densité ont permis le meilleur recouvrement des cellules bactériennes (Lindahl, 1996).

Le recouvrement d'une information exhaustive est cependant dépendant de cette première étape de séparation des cellules de la matrice sol. Aakra *et al.* n'ont par exemple pu extraire que 0,5 % des bactéries oxydant l'ammonium (Aakra *et al.*, 2000). Plus généralement, moins de 50 % de la diversité bactérienne sont extraits du sol (Bakken, 1985; Robe *et al.*, 2003; Steffan *et al.*, 1988) pouvant atteindre exceptionnellement plus de 70 % (Bakken, 1985; Barra Caracciolo *et al.*, 2005). Ces différences peuvent être expliquées par le pourcentage d'argiles présent dans le sol (Bakken, 1985). L'auteur a en effet observé une sédimentation des cellules avec les argiles. Ces dernières fermement attachées aux particules de sol ont une densité plus élevée que les bactéries, ne permettant pas leur isolement sur gradient de densité. L'utilisation de traitements enzymatiques attaquant la matrice EPS des cellules bactériennes permettent une amélioration des rendements d'extraction (Bockelmann *et al.*, 2003). L'extraction à partir de sol non séché est également recommandé, la dessiccation conduisant les cellules à un rapprochement de surface (malgré les répulsions électrostatiques) et donc à un attachement plus fort (Berry *et al.*, 2003).

Les résultats présentés dans la littérature indiquent que l'ADN extrait selon cette voie est de meilleure qualité comparativement à la voie directe. Bien que les acides humiques sédimentent avec les cellules lors de l'étape de centrifugation différentielle (Bakken, 1985; Steffan *et al.*, 1988), ils sont en grande majorité éliminés lors de l'isolement des cellules sur gradient de densité (Bakken, 1985; Lindahl, 1996). Le meilleur niveau de pureté de l'ADN obtenu par cette voie d'extraction facilite son exploitation ultérieure. Ainsi, 2 à 3 fois plus de transformants sont obtenus avec L'ADN extrait par la voie indirecte par rapport à celui extrait par voie directe (Gabor *et al.*, 2003).

La taille des fragments ADN extraits par voie indirecte est également bien supérieure à celle obtenue par la méthode directe. Des fragments de plus de 150 kb (Beja *et al.*, 2000) à plus de 1 000 kb (Berry *et al.*, 2003) ont pu être détectés. Les traitements enzymatiques (lysozyme +/- achromopeptidase) et chimiques (SDS) généralement effectués pour la lyse cellulaire *ex situ* (Courtois *et al.*, 2001; Steffan *et al.*, 1988) sont en effet beaucoup moins sévères que les traitements physiques apportés lors de la lyse *in situ*. Ce traitement lytique enzymatico-chimique ménagé reste néanmoins efficace puisque plus de 90 % des bactéries sont lysées, y compris lorsque le mélange comprend plus de 44 % de bactéries à Gram positif (Steffan *et al.*, 1988). La lyse cellulaire effectuée directement dans des gels d'agarose minimise encore les risques de dégradation mécanique de l'ADN (Beja *et al.*, 2000; Berry *et al.*, 2003; Stein *et al.*, 1996). L'ADN est ensuite séparé du lysat par électrophorèse en champ pulsé, puis extrait du gel par l'emploi de gélase (Beja *et al.*, 2000; Berry *et al.*, 2003) ou par

electroélution (Beja *et al.*, 2000), limitant ainsi la manipulation de l'ADN. Un autre intérêt majeur de l'approche indirecte est l'accès à l'ADN intracellulaire des bactéries, excluant l'ADN extracellulaire procaryotique persistant dans le sol mais également l'ADN eucaryotique.

En conclusion, la voie d'extraction indirecte ainsi que celle directe permet l'accès jusqu'à 60-70 % de l'information génétique de la communauté microbienne contenue dans l'échantillon étudié. Le facteur limitant dans les deux cas provient de l'interaction de l'ADN ou des cellules avec les composants du sol. Les cellules bactériennes montrant une interaction variable avec les particules de sol suivant les espèces (Stenstrom, 1989; van Loosdrecht *et al.*, 1987), l'accès à une diversité biaisée vers les espèces bactériennes moins fortement attachées peut être supposé. Quelques études comparent l'influence de la méthode d'extraction d'ADN métagénomique sur la diversité bactérienne extraite. Les données obtenues sont cependant contradictoires suivant les études. Pour Courtois *et al.* (2001), les résultats observés sur des populations cibles appartenant à trois sous classes de protéobactéries, aux actinomycètes et aux bactéries à faible pourcentage GC sont globalement comparables. Gabor *et al.* (2003) observent une plus grande diversité par la voie d'extraction indirecte (lyse cellulaire *ex situ*). D'autres auteurs montrent une plus forte diversité obtenue par la méthode d'extraction directe (Luna *et al.*, 2006). Néanmoins très peu de taxa révélés sont communs aux deux méthodes, chacune d'entre elles offrant ainsi l'accès à une fraction différente de la microflore tellurique (Luna *et al.*, 2006).

## **5 EXPLOITATION DU METAGENOME**

Les approches moléculaires peuvent permettre l'analyse jusqu'à près de 50-70 % de la diversité bactérienne présente dans un échantillon de sol (Barra Caracciolo *et al.*, 2005), soit 160-230 fois plus que par les approches classiques d'isolement bactérien *in vitro* (Amann *et al.*, 1995). Une telle quantité d'information nécessite des approches adaptées pour leur exploitation. Plusieurs approches ont ainsi été développées (**Figure 1**) pour un contexte d'étude particulier.

## 5.1 Analyse de la communauté bactérienne via l'outil PCR

Parmi les approches développées pour étudier la diversité bactérienne, certaines utilisent la technique PCR pour une analyse ciblée de la diversité bactérienne (Kirk *et al.*, 2004; Ranjard *et al.*, 2000). La définition d'un couple d'amorce dans les régions conservées du gène étudié va permettre l'amplification ciblée de ce gène tout en accédant à la diversité des séquences nucléotidiques encadrée par ces régions. L'utilisation de séquences composites (dites dégénérées), et des conditions d'amplification adaptées, peuvent augmenter la diversité des séquences alors amplifiées par l'admission de séquences plus divergentes (Metsa-Ketela *et al.*, 2002; Philippot *et al.*, 2002).

Des gènes à critères taxonomiques tel que le gène *rrs* (codant la sous unité ribosomique 16S), l'espace intergénique *rrs-rrl*, ou des gènes dits fonctionnels sont ainsi amplifiés à partir d'ADN métagénomique pour l'analyse de leur diversité dans l'échantillon étudié. Des amorces ont notamment été définies pour les gènes *nifH* (Gros *et al.*, 2006; Hery *et al.*, 2005b) et *narG* (Hery *et al.*, 2005b; Philippot *et al.*, 2002) ciblant les populations fixatrices d'azotes et les populations dénitrifiantes, ainsi que pour d'autres gènes intervenant dans la production de composés antibiotiques tels que *phlD* (Bergsma-Vlami *et al.*, 2005) ciblant les populations productrices de 2,4-diacétylphloroglucinol, et les gènes *pks* (Wawrik *et al.*, 2005) ciblant les populations productrices de polykétides.

Cependant certains biais inhérents à la PCR contraignent les études basées sur ces produits d'amplification. L'amplification de produits chimériques peut conduire à une surestimation de la diversité bactérienne. Des différences d'affinités des amorces pour l'ADN cible, ainsi que le pourcentage GC de l'ADN cible peuvent amener à l'amplification préférentielle de certaines espèces. Par ailleurs les effets de dilution de l'ADN cible peuvent entraîner l'amplification préférentielle des populations dominantes (von Wintzingerode *et al.*, 1997).

Ce type d'approche se révèle cependant intéressante pour la comparaison de profils entre différentes situations environnementales. L'importance des biais PCR est ainsi limitée puisque les analyses effectuées sont relatives à un profil donné. Des produits d'amplification du gène exprimant l'ARNr16S ont ainsi été clonés puis séquencés afin d'identifier les modifications majeures de diversité taxonomique liées à l'apport de nutriments dans un sol pauvre en carbone (Hery *et al.*, 2005a). D'autres voies d'analyse font intervenir des techniques plus spécifiques pour l'analyse des modifications de communautés. Les produits d'amplification ont par exemple été analysés par T-RFLP pour observer les modifications de

diversité entre des agrégats de différentes tailles (Mummey & Stahl, 2004), ou par hybridation sur des puces à ADN pour observer l'impact rhizosphérique sur les communautés microbiennes (Sanguin *et al.*, 2006).

## 5.2 Séquençage direct de l'ADN métagénomique

D'autres approches développées visent au séquençage direct de l'ADN métagénomique. Une voie de séquençage validée pour la reconstruction du génome de *Haemophilus influenzae* (Fleischmann *et al.*, 1995) a été adaptée et appliquée sur le métagénome extrait des environnements étudiés. L'ADN est fragmenté en petits morceaux de 2 kb, cloné puis séquencé. Les séquences nucléotidiques obtenues sont ensuite assemblées par analyse bioinformatique.

Cette méthode, appliquée sur l'ADN extrait d'eaux de drainage de mine de fer, a permis la reconstruction du génome quasi complet de 2 bactéries appartenant aux espèces *Leptospirillum sp.* et *Ferroplasma sp.* ainsi que le génome partiel de 3 autres bactéries (Tyson *et al.*, 2004). Cette reconstruction génomique a été rendue possible par la faible diversité de l'environnement étudié. La reconstruction de génomes de toutes les espèces bactériennes présents dans l'environnement nécessite cependant le séquençage d'au minimum 3 fois leur couverture génomique, voire beaucoup plus suivant la richesse des espèces et leur abondance (Fleischmann *et al.*, 1995; Tyson *et al.*, 2004; Venter *et al.*, 2004).

L'utilisation la plus spectaculaire de cette approche a été le séquençage systématique de l'ADN extrait d'échantillons de la mer des sargasses (Venter *et al.*, 2004). Plus de  $1 \times 10^{12}$  pb ont été traités et assemblés, identifiant ainsi près de 1,2 millions de gènes inconnus. Une telle approche a permis l'identification de 148 phylotypes inconnus basée sur la similarité de séquence (seuil de 97 %) des gènes exprimant l'ARNr 16S.

Une telle approche permet ainsi le dénombrement des différentes espèces ainsi que les différentes fonctions présentes dans l'environnement étudié, sans les biais liés à la PCR (Eisen, 2007). Les données ainsi collectées sont sources d'analyses plus globales sur la comparaison d'écosystèmes (Tringe *et al.*, 2005) ou sur la diversité et l'évolution des gènes (Kannan *et al.*, 2007; Yooseph *et al.*, 2007).

### 5.3 Construction et exploitation des banques métagénomiques

Une autre stratégie d'exploitation de l'ADN métagénomique implique le clonage de fragments d'ADN de plus ou moins grande taille dans une souche réceptrice pour confectionner des banques dites métagénomiques. Les objectifs de cette banque vont au-delà du séquençage systématique de la totalité de l'ADN cloné, par une analyse ciblée sur des gènes à critères taxonomiques ou fonctionnels. La banque construite est alors criblée selon deux voies pour l'identification des clones portant l'information génétique ciblée.

Une première voie de criblage dite moléculaire permet l'identification des clones selon la séquence nucléotidique du fragment métagénomique cloné. Ainsi l'approche PCR ou l'hybridation avec une sonde permet la détection de ces clones par similarité de séquence (Daniel, 2004).

Une deuxième voie de criblage est basée sur l'expression des gènes présents sur l'insert métagénomique. Une telle approche permet d'identifier les fonctions sans a priori sur la séquence nucléotidique. Ce point est particulièrement intéressant dans la recherche appliquée pour la découverte de nouveaux biocatalyseurs ou de nouvelles activités antibactériennes (Daniel, 2004). Bien que l'expression nécessite la présence chez la souche hôte du matériel génétique nécessaire à la transcription / maturation de la protéine (facteurs de transcription, inducteurs, cofacteurs, protéines chaperones, etc) (Gabor *et al.*, 2004a; Gabor *et al.*, 2004b), une expression hétérologue a pu être observée chez *E. coli*. De nombreuses activités enzymatiques exprimées par des clones ont en effet été détectées, et de nouvelles molécules ont pu être isolées (Brady *et al.*, 2002; Courtois *et al.*, 2003; Rondon *et al.*, 2000; Wang *et al.*, 2000). Des vecteurs navettes ont également été développés pour faciliter leurs transferts de *E. coli* vers différents hôtes (Martinez *et al.*, 2004; Schloss & Handelsman, 2003; Sosio *et al.*, 2000), augmentant ainsi les probabilités d'expression des ADN métagénomiques.

Deux stratégies de clonage ont été adoptées. La première opte pour le clonage de petits fragments d'ADN (moins de 15 kb) dans des plasmides (Gabor *et al.*, 2004b). Le grand nombre de copies du plasmide améliore la détection d'activités faiblement exprimées, et favorise l'expression des gènes clonés sous le contrôle de promoteurs apportés par le vecteur (Daniel, 2005), bien que la probabilité d'expression d'ORFs (*open reading frame*) dans cette configuration reste faible (Gabor *et al.*, 2004a). La deuxième stratégie vise au clonage de grands fragments (40 à 100 kb) dans des vecteurs à faible nombre de copies tels que les cosmides, fosmides ou les BACs (*bacterial artificial chromosome*) et l'expression de l'ADN sous le promoteur propre des gènes clonés. Par cette approche, le clonage d'opérons ou de



voies biosynthétiques complètes peut être projeté (Daniel, 2005). Le risque est cependant plus élevé que les clones portant des gènes toxiques soient exprimés (Fleischmann *et al.*, 1995), entraînant la mort de la souche hôte et donc la perte d'une partie des informations génétiques de la microflore tellurique.

L'ADN extrait selon les deux voies d'extractions, directe et indirecte, pourraient donc dans les deux cas être exploités pour la construction de banques métagénomiques. L'ADN métagénomique extrait par la méthode directe, générant des petits fragments, pourrait être utilisé pour la construction de banques métagénomiques présentant des inserts de petite taille, alors que celui extrait par la méthode indirecte permettra la construction de banque métagénomique portant des grands inserts. Cependant, plusieurs points conduisent à préférer la voie d'extraction indirecte (lyse cellulaire *ex situ*) pour la construction de banques métagénomiques. Premièrement, l'ADN extrait par la voie directe est plus contaminé par les acides humiques que l'ADN extrait selon la voie indirecte. Ces acides humiques, interférant avec les réactions enzymatiques (Wilson, 1997) et la transformation d'ADN (Gabor *et al.*, 2003), vont mener à une perte plus importante d'information génétique lors de son clonage. Cet ADN est également contaminé par l'ADN extracellulaire présent dans le sol, dont l'ADN eucaryotique (Gabor *et al.*, 2003). Les banques construites contiendraient ainsi majoritairement de l'ADN eucaryotique. Pour finir, la construction d'ADN métagénomique à partir de grands fragments ADN contribue à diminuer la taille des banques à construire nécessaire pour recouvrer la diversité bactérienne présente dans l'échantillon étudié.

De façon assez générale, la méthode indirecte est privilégiée pour la construction de banques métagénomiques. L'ADN extrait selon la voie directe se prête mieux à une analyse par amplification PCR où une information génétique ciblée est étudiée. Rappelons cependant que d'après Luna *et al.* (2006) des informations génétiques différentes sont contenues dans l'ADN métagénomique selon la voie d'extraction (directe ou indirecte) utilisée. Selon le contexte d'étude, différentes voies d'extraction et d'exploitation de l'ADN métagénomique sera donc utilisée. Dans le cas où une information exhaustive est recherchée, une combinaison des approches pourrait se révéler nécessaire.

## **6 CONCLUSION**

Le sol est un environnement structuré, dans lequel sont distribués de manière hétérogène les microorganismes. Les propriétés physico-chimiques du sol affectent le recouvrement de l'information génétique des bactéries telluriques présentes dans l'échantillon étudié, soit par la difficulté d'extraction des cellules de la matrice, soit par les problèmes de désorption et de purification de l'ADN des constituants du sol. Dans le meilleur des cas, seulement 60 à 70 % de l'information génétique serait extraite. Une extraction exhaustive des informations génétiques contenues dans matrice sol apparaît donc particulièrement ardu.

Les informations collectées ont permis de nombreuses avancées tant au niveau fondamental qu'au niveau appliqué. Néanmoins, la recherche d'une information exhaustive implique une stratégie adaptée pour recouvrir l'ensemble de la diversité bactérienne tellurique. Chacune des stratégies d'extraction présentées ci-avant donnant accès à une information génétique différente, la combinaison des deux stratégies permettraient par exemple d'accéder à une représentation plus complète de la diversité. Les biais de chacune des méthodes présentées ci-avant pourraient ainsi être compensées par la multiplication des approches afin d'obtenir une image plus complète de la diversité bactérienne.

Toutefois, une nouvelle approche évitant à la fois les biais d'isolement en culture pure et les biais liés à l'extraction de l'ADN bactérien de la matrice sol permettrait de compléter les informations génétiques jusqu'à présent collectées.



## **CHAPITRE II : RESULTATS**

### **FACTEURS INFLUENCANT LA QUALITE DE L'ADN BACTERIEN EXTRAIT DU SOL**

La constitution et l'utilisation de banques d'ADN métagénomique constituent l'approche la plus prometteuse pour exploiter à des fins fondamentales comme appliquées les informations de l'ADN bactérien extrait de l'environnement. Toutefois, le clonage d'une très importante quantité d'ADN sera requis dans les cas où les objectifs sont de créer une banque d'ADN représentative du niveau très élevé de la diversité bactérienne dans un environnement complexe comme le sol.

Le nombre de clones dans la banque d'ADN ainsi que la taille moyenne des inserts sont les deux facteurs qui permettent d'évaluer et potentiellement d'ajuster la quantité d'ADN métagénomique cloné, en considérant que plus de 9 millions de clones avec des inserts en moyenne de 40 kb sont nécessaires pour couvrir 1 seule fois un métagénome constitué de l'ADN provenant de 10 000 espèces génomiques différentes (Lombard *et al.*, 2006). Les inserts de grande taille présentent de nombreux intérêts outre le fait qu'ils permettent de diminuer le nombre de clones à générer, manipuler et cribler. La présence d'une longue séquence nucléotidique sur l'insert accroît la probabilité d'y localiser des gènes d'intérêt taxonomique (dont les gènes ribosomiques) permettant d'identifier l'hôte initial du fragment d'ADN cloné (Beja *et al.*, 2000; Liles *et al.*, 2003; Treusch *et al.*, 2004). Ce n'est enfin qu'à partir de telles stratégies de clonage de grands inserts que seront isolés des opérons complets ou que pourra être développée une technologie de criblage chimique de molécules produites dans le surnageant de culture après expression en système hétérologue de ces gènes inconnus (Courtois *et al.*, 2003; Entcheva *et al.*, 2001; MacNeil *et al.*, 2001).

Les objectifs des travaux qui sont présentés dans ce chapitre et qui ont été publiés dans la revue Journal of Microbiological Methods (Bertrand *et al.*, 2005) ont été d'optimiser les technologies d'extraction de l'ADN bactérien à partir de deux sols en vue de développer les protocoles protégeant le plus possible l'intégrité de l'ADN. Ces objectifs nous ont conduits à étudier l'influence de l'état physiologique des bactéries *in situ* sur la qualité de l'ADN extrait et à tenter de modifier ces paramètres physiologiques bactériens par ajout de nutriments dans

le sol afin de vérifier l'implication de tels traitements tant sur l'état de l'ADN extrait que sur l'évolution de la communauté bactérienne.

## **HIGH MOLECULAR WEIGHT DNA RECOVERY FROM SOILS PREREQUISITE FOR BIOTECHNOLOGICAL METAGENOMIC LIBRARY CONTRUCTION**

Hélène Bertrand, Franck Poly, Van Tran Van, **Nathalie Lombard**, Renaud Nalin, Timothy M. Vogel, Pascal Simonet

### **Abstract**

Soil is a complex environment considered as one of the main reservoirs of microbial diversity. However, the inability to cultivate most soil bacteria hampered fundamental attempts to determine the diversity of the prokaryotic world and limited its industrial exploitation. In the last 20 years, new methods have been developed to overcome these limitations based on the direct extraction of DNA from bacteria in their natural environment. In addition to fundamental research, the cloning of the extracted DNA for the development of metagenomic DNA clone libraries offers possibilities to discover novel bio-molecules through the expression of genes from uncultivated bacteria in surrogate bacterial hosts. However, such objectives require adapting DNA extraction methods and cloning strategies in order that entire gene clusters encoding biosynthetic pathway for secondary metabolites can be cloned. In this paper, we report that the size of DNA fragments extracted from soil varied in a range between less than 100 kb to more than 400 kb depending on the soil. The relatively limited size of DNA fragments extracted from some soil was not only due to mechanical, chemical or enzymatic shearing of the DNA during the extraction process but partly to the microbial growth status. Stimulating bacteria *in situ* by providing nutrients to the soil improved the size of extracted DNA, but it modified the bacterial community structure.

## 1 **INTRODUCTION**

An impressive accumulation of physiological, genetic and ecological data have been obtained for some microorganisms including whole genome sequences of about 200 bacteria and Archaea. However, some bacterial phyla remain totally unknown so that the actual extent of microbial diversity is still impossible to define (Rossello-Mora and Amann, 2001). This gap is due to the difficulty in cultivating many bacteria, and thus, reducing the applications of the traditional investigation methods that are based on *in vitro* clonal multiplication.

Efforts made to develop new culture media based on a better understanding of bacterial requirements will inevitably increase the number of available isolates (Joseph et al., 2003), but access to the majority of microbial diversity will remain difficult. Alternate strategies for studying bacterial diversity do not rely on *in vitro* bacterial cultures, but are based on the recovery of DNA from entire microbial communities. Such an approach, called metagenomic, has become commonplace when combined with one of numerous PCR derivative techniques. This approach can gather taxonomy and phylogeny related information when targeting ribosomal or other conserved genes (Torsvik and Ovreas, 2002). In addition to accessing community phylogenetic diversity, metagenomic DNA also contains functional genes from uncultivated bacteria and, hence, constitutes a promising source of novel metabolites (Daniel, 2004). Potential discovery of these compounds requires the metagenomic DNA to be cloned and the resulting metagenomic libraries screened adequately in order to detect new compounds produced in surrogate hosts. Preliminary analysis of some cloned inserts from the first soil metagenomic libraries validated the feasibility and interest of this approach (Daniel, 2004). The level of DNA diversity and novelty in metagenomic library inserts corresponds to that expected on the basis that 99 % of soil bacteria are uncultivated (Courtois et al., 2003). Enzymes and other compounds that were already produced by some recombinant clones confirmed the possibility for heterologous expression of foreign DNA in new hosts (Seow et al., 1997; Wang et al., 2000 ; MacNeil et al., 2001 ; Daniel, 2004).

However, two parameters have to be carefully considered when developing metagenomic libraries. The first one is the large size of the metagenome, which is defined as the combination of all genomes from bacteria colonizing a given environment. According to reassociation kinetic data, the genetic diversity of the soil metagenome would be between 5,000 and 500,000 times that of the *E. coli* genome (Torsvik et al., 1998; Doolittle, 1999). Such diversity requires improved cloning efficiency so that the clones in the gene library provide an acceptable representation of the entire metagenome. The second consideration is

**Table 1: Châteaurenard and Montrond soil characteristics (g.kg<sup>-1</sup>).**

Soil	Clays (< 2 µm)	Loam (2-20 µm)	Coarse silts (20-50 µm)	Fine sands (50-200 µm)	Coarse sands (200-2000 µm)	
CR <sup>a</sup>	151	246	237	309	57	
MR <sup>b</sup>	182	251	131	172	264	
Soil	Organic C	O M <sup>c</sup>	pH water	pH KCl	EC <sup>d</sup>	
CR <sup>a</sup>	10.5	18.0	8.2	7.75	59	
MR <sup>b</sup>	27.3	47	5.9	5.38	139	
Soil	Total CaCO <sub>3</sub>	Calcium	Magnesium	Potassium	Phosphorus	Fe oxalate
CR <sup>a</sup>	423	7.7	0.192	0.218	0.524	0.027
MR <sup>b</sup>	0	2.84	0.220	0.379	0.506	1.138

<sup>a</sup>CR Châteaurenard,

<sup>b</sup>MR Montrond

<sup>c</sup> Organic Matter

<sup>d</sup> Cationic exchange capacity (meq.kg<sup>-1</sup>)

the size and the cluster organization of genes involved in the synthesis of secondary metabolites. The typical size of gene clusters requires cloning relatively large DNA fragments in order to improve the probability of isolating an entire biosynthetic pathway.

The first soil DNA extraction techniques (direct methods) were based on the *in situ* lysis of bacteria in soil prior to DNA recovery and purification (Ogram et al., 1987 ; Picard et al., 1992). However, in order to limit mechanical shearing of the DNA, contact between DNA and soil components, and DNA degradation, a second strategy (indirect methods) was developed in which bacterial cells are extracted from the soil matrix prior to cell lysing and DNA purification. This second approach (indirect methods) typically produces longer DNA fragments. In addition, the extracted bacterial cell pellet can be submitted to nuclease and other treatments that could increase cloning efficiency and limit contamination by eukaryotic extracellular DNA (Gabor et al., 2003).

The aim of this work was to investigate the link between soil microbiological status and the DNA size limit obtained by the DNA extraction method. Two soils differing in their traditional use (grassland soil from Montrond and long term cultivated soil from Châteaurenard, France) were used as sources for metagenomic DNA. An “indirect extraction” method was developed to limit mechanical, enzymatic and chemical DNA degradation. In addition, soil enrichment steps prior to cell extraction were tested for their influence on DNA quality and quantity and also for the diversity bias they might induce in the recovered metagenomic community structure.

## **2 MATERIALS AND METHODS**

### **2.1 Soil samples**

Soil from Châteaurenard (Bouches du Rhône, France) and from Montrond (Isère, France) were collected from the upper 5 to 20 cm. After root removal, soil samples were sieved through a 2 mm mesh and stored at 4°C. Their compositions are presented in **Table 1**.

### **2.2 Extraction of soil bacteria**

The extraction of soil bacteria was performed on the two fresh soil samples using a Nycodenz gradient separation method. The equivalent of 100 g of dry soil was homogenized in a Waring blender with 150 mL sodium hexametaphosphate solution 0.2 % for 1 min. Coarse soil particles were eliminated through a first centrifugation step at low speed (700 x g,



15 min, 10°C in a Beckman J2-21M/E centrifuge). The supernatant was collected, filtered with gauze and centrifuged at 7500 x g during 20 min at 10°C. The microbial cell fraction-containing pellet was resuspended in sterile 0.8 % sodium chloride solution by vortexing to obtain a homogenous suspension. Twenty-five milliliters of the suspension were transferred to an ultracentrifuge tube containing 10 mL of Nycodenz solution (Axis-Shield, Oslo, Norway) at 1.3 g mL<sup>-1</sup> density (8 g of Nycodenz to 10 mL of sterile ultra pure water). Cells and soil particles were separated in a second centrifugation step at high speed (14600 x g, 40 min at 10°C) in a Kontron TST 2838 swing-out rotor with a Kontron TGA-50 ultracentrifuge. A white band of bacterial cells was obtained at the interface between the Nycodenz-soil mix particles and the overlying aqueous layer. This band was carefully recovered; cells were resuspended in sterile ultra pure water and centrifuged twice at 7500 x g (20 min, 10°C) to remove the Nycodenz solution.

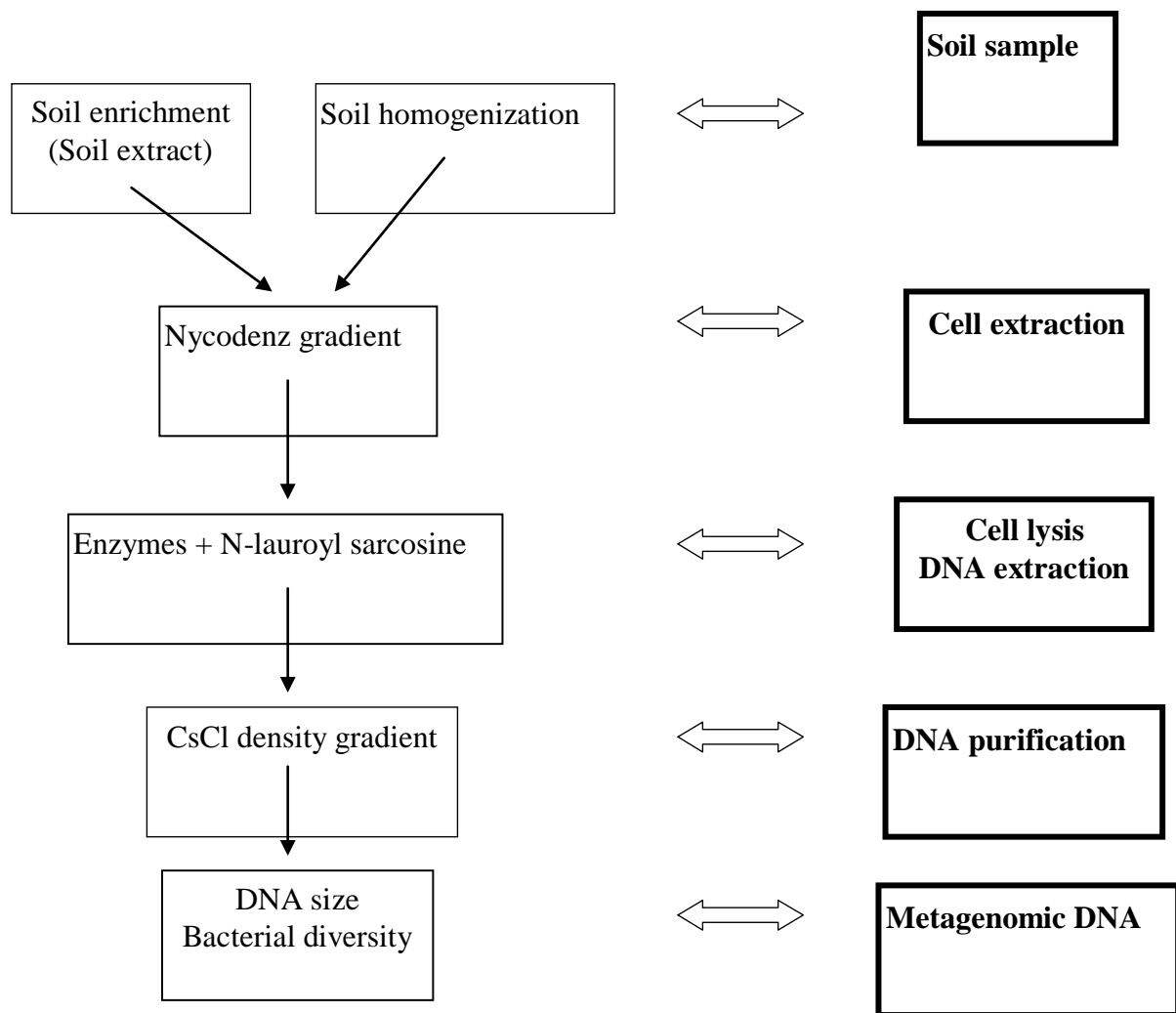
### **2.3 Bacterial counts**

Ten grams (equivalent dry weight) of the two soils were mixed ground with 10 mL of sodium hexametaphosphate solution 0.2 % for 1 min in the Warring blender and soil suspensions were serially diluted. Serial dilutions were also performed with bacterial cells obtained from the Nycodenz gradient. The bacterial number present in soil directly or after extraction was determined by acridine orange count (Richaume et al., 1993)

### **2.4 Enrichment cultures**

Soil enrichments used soil extracts (Pochon and Tardieux, 1962) prepared as described in the following protocol. One kg of soil was mixed with 1 L of deionised water and subsequently autoclaved at 130°C for 1 hour. This suspension was decanted overnight at 4°C. The upper phase was centrifuged at 700 x g for 10 min and then at 7500 x g for 30 min to remove small particles. The pH was adjusted to 6.8. This soil extract was then autoclaved at 115°C for 30 min.

The enrichments were performed with 200 g of Châteaurenard fresh soil (equivalent to 185 g of dry soil) and 13 g of Montrond fresh soil (equivalent to 10 g of dry soil) to which 800 mL and 100 mL, respectively of the corresponding soil extract were added. These enrichments were incubated at 28°C and shaken at 150 rpm for 60 h before separation of cells and soil particles with Nycodenz gradient separation.



**Fig. 1: Protocol of indirect bacterial DNA extraction and purification from soil**

## 2.5 Cell lysis and DNA purification

The bacterial pellets obtained from soil or enriched soil were resuspended in 2 mL of Tris-EDTA (TE) (10 mM Tris-HCl, 100 mM EDTA). The cells were lysed by adding 2.5 mL of lyse buffer (TE, 5 mg mL<sup>-1</sup> lysozyme, 0.5 mg mL<sup>-1</sup> achromopeptidase) during incubation for 1 h at 37°C (Courtois et al., 2001). Then, proteinase K (1 mg mL<sup>-1</sup>) and N-lauroyl sarcosine (1 %) were added for 45 min at 50°C. The DNA was purified in a cesium chloride density gradient without ethidium bromide: 6.83 g cesium chloride and 1.88 mL of TE were added to the lysed cell suspension and centrifuged with a TFT 6513 Kontron rotor for at least 39 h at 10000 x g and 15°C. The DNA was then dialyzed in a TE (10 mM Tris-HCl, 1 mM EDTA) buffer to eliminate the cesium chloride. DNA amounts were determined at 260 nm using an UV spectrophotometer (Biophotometer Eppendorf). The DNA obtained was analyzed in order to determine (i) the fragment size (ii) and the bacterial diversity. The method is summarized in **Fig. 1**.

## 2.6 DNA fragment size determination by pulsed field electrophoresis

The DNA fragment sizes were determined by pulsed field electrophoresis. The DNA obtained by purification on cesium chloride density gradients was placed in 0.8 % or 1 % standard agarose gel and the migration was performed in 0,5 X Tris Borate EDTA (TBE) buffer in a Bio-Rad CHEF-DRII apparatus at 14°C. The migration parameters, such as time, pulse time and voltage, depended on the DNA fragment sizes. The *lambda* ladder PFG marker and the low range PFG marker were used (BioLabs).

## 2.7 Direct extraction and purification of DNA

The DNA from 500 mg of the two soils studied was directly extracted and purified with the FastDNA SPIN Kit for soil (BIO 101, Qbiogen). DNA amounts were determined at 260 nm using an UV spectrophotometer (Biophotometer Eppendorf).

## 2.8 Structure of soil microbial community determined by ribosomal intergenic spacer analysis (RISA)

The intergenic spacers between the small- and large-subunit rRNA genes were amplified using the primers F72-16S (eubacterial rRNA small subunit, 5'-TGCGGCTGGATCCCCCTCCTT-3') and F38-23S (eubacterial rRNA large subunit, 5'-CCGGGTTTCCCCATTCGG-3') (Normand et al., 1996). Amplified sequences contained the Intergenic Spacer (ITS). Reaction mixtures (50 µL) for the PCR contained 5 µL of 10 X

dilution buffer Invitrogen 250 ng T4 gene 32 protein (Qbiogen), 0.5  $\mu$ M of each primer, 200  $\mu$ M of each dNTP, 2.5 U of Taq DNA polymerase (Invitrogen) and 10 ng of purified soil DNA. Amplification was performed at 94°C for 3 min, followed by 25 cycles consisting of 94°C for 1 min, 55°C for 1 min, and 72°C for 1 min, and extension of incomplete products for 5 min at 72°C.

Fragments ranging in size from approximately 300 to 1100 bp were resolved on 5 % (w/v) polyacrylamide (37,5:1) –Tris– borate–EDTA gels (1.5 mm thick, 30 cm long). Gels were run at constant temperature (15°C) for 16 h at 65 V and 5 mA (DSG200-02, C.B.S. Scientific, Del Mar, Ca) and then stained with SYBR Green I (FMC Bioproducts, France) according to the manufacturer's instructions. GelCompar software (Applied Maths, Kortrijk, Belgium) was used to convert fingerprint profiles to numeric data. A statistical analysis of these data profiles was performed by principal components analysis (PCA) (Poly et al., 2002) using ADE-4 software (Thioulouse et al., 1997).

The structure of the soil microbial community was determined with DNA extracted and purified by both direct and indirect methods.

### **3 RESULTS AND DISCUSSION**

#### **3.1 Extraction and purification of soil DNA**

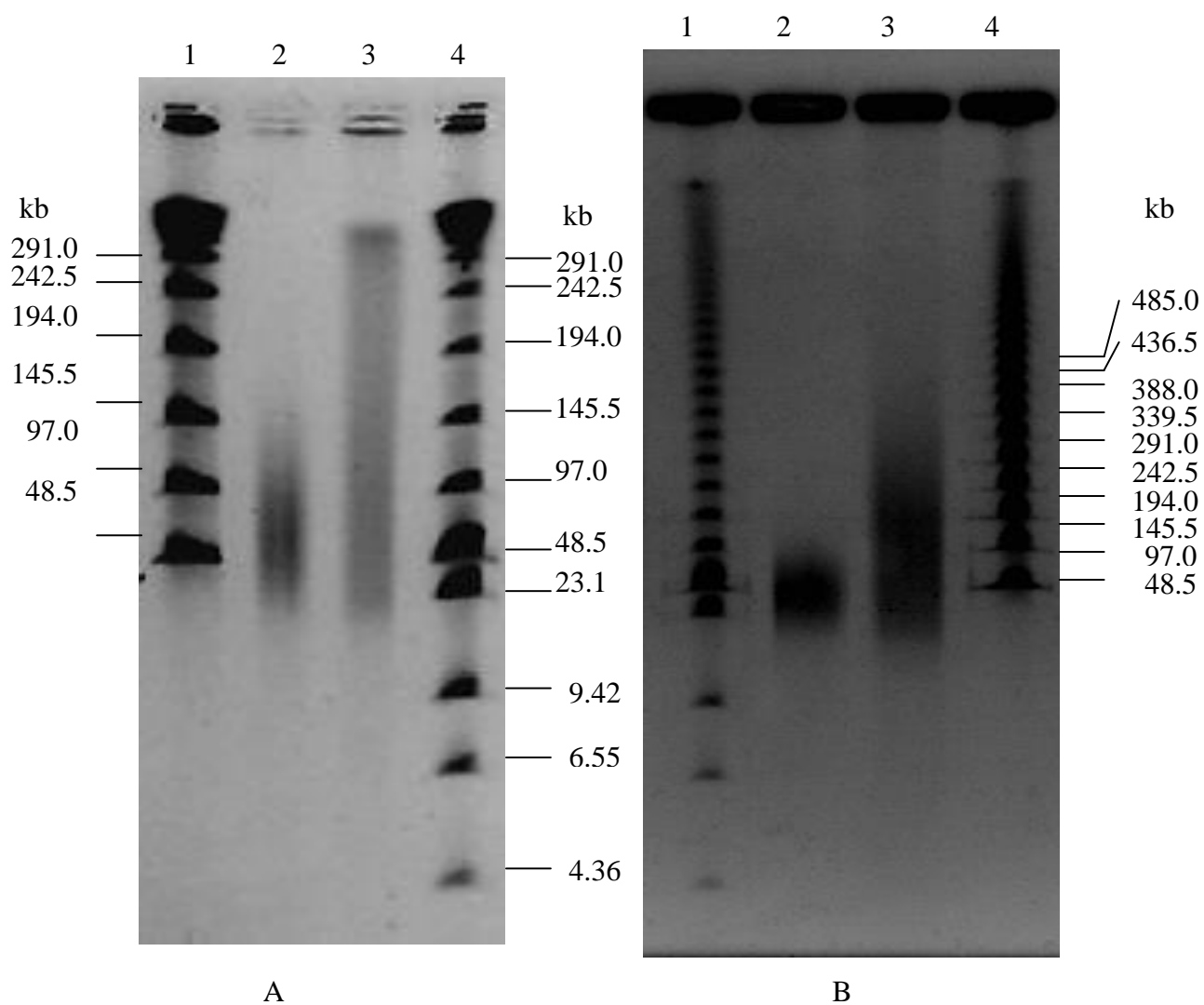
Generating soil metagenomic libraries with clones containing long DNA inserts requires developing and using adapted methods for the extraction and purification of undegraded DNA from soil bacteria. The requirement for undegraded DNA excludes the use of the DNA-shearing “direct” extraction method in which lysis treatments are carried out *in situ* before DNA extraction and purification (Robe et al., 2003). The “indirect” method appears to be better adapted to extracting undegraded DNA because the extraction of the bacterial pellet in the Nycodenz gradient limited contact between the DNA and soil components. In addition, the bacterial fraction was lysed less vigorously than during the direct method with equal success.

The two soils used in this study were both classified as sandy clayey silt soils but the organic matter content in Montrond soil was 2.6-fold greater than that in Châteaurenard soil (Table 1). A direct estimation of bacterial biomass by the acridine orange staining method indicated that the two soils also differed by the number of bacteria with Montrond soil having 6-fold more bacteria than the Châteaurenard soil ( $6 \times 10^9 \pm 1.1 \times 10^9$  and  $1 \times 10^9 \pm 1.9 \times 10^8$  cell.g<sup>-1</sup>

dry soil respectively). The proportion of bacteria extracted from the two soils by the Nycodenz method was similar to the ratio of the bacterial numbers for the two soils ( $2.2 \times 10^8 \pm 6.9 \times 10^6$  and  $3.5 \times 10^7 \pm 1.6 \times 10^5$  cell number.g<sup>-1</sup> dry soil for Montrond and Châteaurenard soils respectively). These two soils harbored a markedly different microbial biomass that could be due in part to their difference in organic matter content. Thus, the soil characteristics did not appear to bias the bacterial extraction yields when using Nycodenz gradient separation. These results differ from those reported by Bakken (1985) who found that the bacteria extraction yield varied over a wide range with the soil type when using a blending centrifugation procedure.

The second step of the 'indirect' method included DNA extraction and purification from the extracted bacteria with the objective of isolating undamaged DNA. Among the protocols that were developed for extracting DNA or isolating large plasmids from bacteria growing *in vitro*, the method based on the *in situ* lysis of bacteria directly in the cesium chloride (CsCl) density gradient tubes prior to ultracentrifugation (Schwinghamer, 1980) appeared to be the most promising for soil extracted bacterial pellets (Courtois et al., 2003). This method excludes any manipulation of extracellular DNA, thus, limiting the risk of physical nicks in the DNA. Moreover, the ultracentrifugation of the CsCl-based isopycnic gradient efficiently produced relatively pure DNA. This treatment recovered 3.75 ( $\pm 0.7$ ) and 0.077 ( $\pm 0.004$ )  $\mu\text{g}$  DNA per gram of dry soil from Montrond and Châteaurenard soils, respectively. These results are consistent with the number of bacterial cells extracted from Châteaurenard soil if the average amount of DNA per bacterial cell is 1.6 fg (Bakken and Olsen, 1989). The estimated DNA yield would be 0.06  $\mu\text{g}$  DNA.g<sup>-1</sup> of soil. The amount of DNA extracted from Montrond soil was 1 order of magnitude greater than that expected (0.35  $\mu\text{g}$  DNA.g<sup>-1</sup>), which not only confirmed the lysis efficiency but also suggested that the bacteria number might have been underestimated and / or that this soil is colonized by bacteria with larger genomes for which the average amount of DNA would be greater than 1.6 fg per cell.

In order to construct a metagenomic library, both the quantity and quality of the DNA are important. The successful enzymatic restriction of DNA confirmed the DNA quality. (data shown).



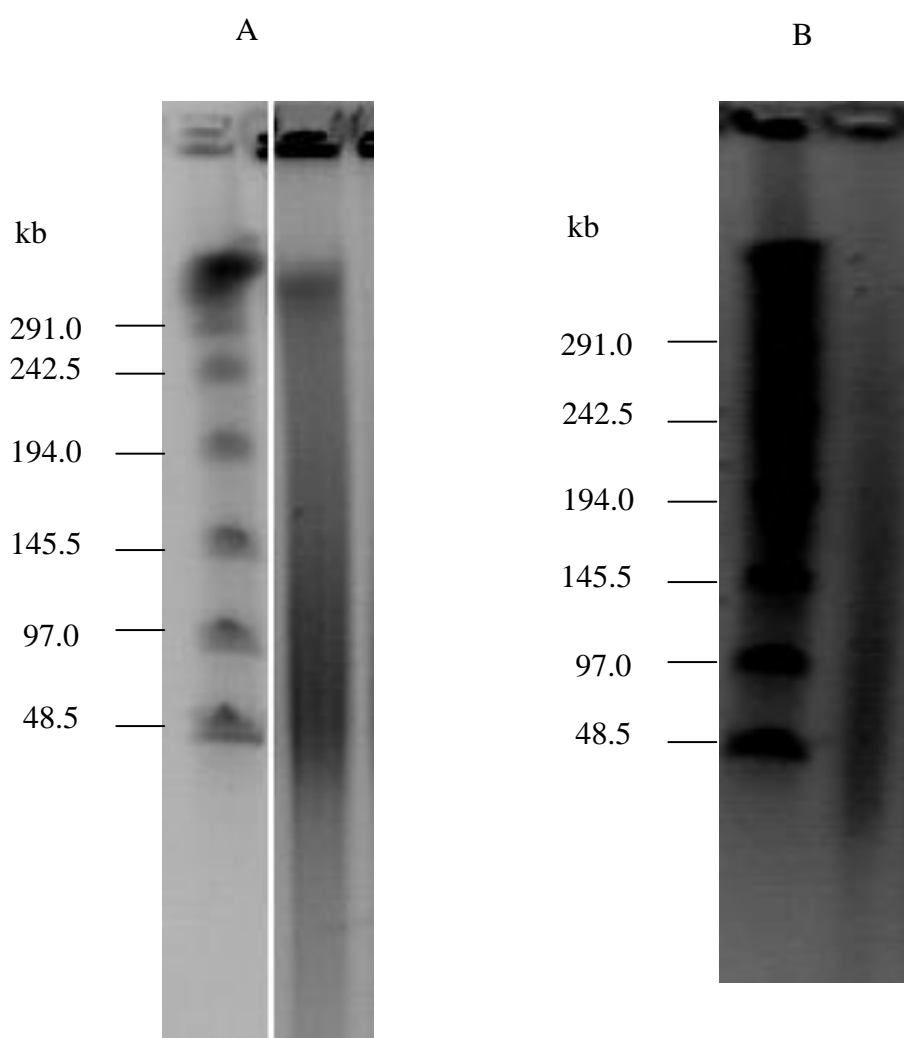
**Fig. 2: PFGE of bacterial soil DNA extracted by indirect method from Montrond soil (Lanes 2) and Châteaurenard soil (Lanes 3). Two amount of soil DNA were loaded 375 ng (A) or 625 ng (B). Lane 1A and 4B: Lambda ladder PFG marker and Lane 4A and 1B: low range PFG marker**

### 3.2 DNA fragment size

Cesium chloride density gradient purified DNA was recovered according to standard procedures, but with a particular care to avoid mechanical shearing. The size of the DNA was estimated by pulsed field gel electrophoresis on 0.8 % agarose gel for 17 h at  $4.5 \text{ V.cm}^{-1}$  with switch times ramped from 10-25 seconds. The two soils provided markedly different results with DNA fragments ranging in size from 25 to 145 kb for DNA extracted from the Montrond soil while fragments up to 300 kb were detected in DNA extracted from the Châteaurenard soil (**Fig. 2A**). Modifying the electrophoresis parameters (21 h runs on a 1 % agarose gel at  $5 \text{ V.cm}^{-1}$  with a switch times ramped from 5-120s) improved detection of DNA fragments from Châteaurenard soil up to 400 kb even if most of the DNA was shorter. For both soils, the DNA smear was almost undetectable in the lower part of the gel (fragment size between 4 and 20 kb) indicating that the DNA solution contained only a low proportion of the small DNA fragments.

Several hypotheses can be proposed to explain the partial degradation of DNA extracted from soil bacteria. Partial DNA degradation could be due to a residual nuclease activity during cesium chloride gradient DNA purification in spite of the high EDTA concentration (100 mM) used in the lytic solution. In order to remove any possibility of residual enzymatic activity, lysis and cesium salt gradient based purification treatments were carried out at  $4^{\circ}\text{C}$  with cesium trifluoroacetate (CsTFA) (Zolg et al., 1988 ; Zhang et al., 2003) replacing cesium chloride (CsCl) to avoid salt crystallization during ultracentrifugation at low temperature. However, this treatment did not lead to a significant increase in the DNA size as evaluated by pulsed field gel electrophoresis. Thus, DNA damage by enzymatic degradation during cell lysis was unlikely.

The two soils differed also in iron concentration with the Châteaurenard soil containing 40 times less iron than the Montrond soil (Table 1). This difference could explain the gap in DNA size between the two soils since ferrous iron (FeIII) chelated to DNA may be involved in DNA damage (Netto et al., 1991) induced by soil oxidizable components. In order to determine if the removal of iron could increase the size of extracted DNA, the Nycodenz gradient extracted cells were washed in a 100 mM KCN solution at room temperature to eliminate FeIII. Cells were centrifuged to remove the KCN solution. This treatment was repeated three times before the pellet was rinsed several times with TE buffer to remove trace KCN. The DNA was then extracted according to the CsCl gradient protocol. However, the patterns of DNA on pulsed field gel electrophoresis did not show any improved DNA



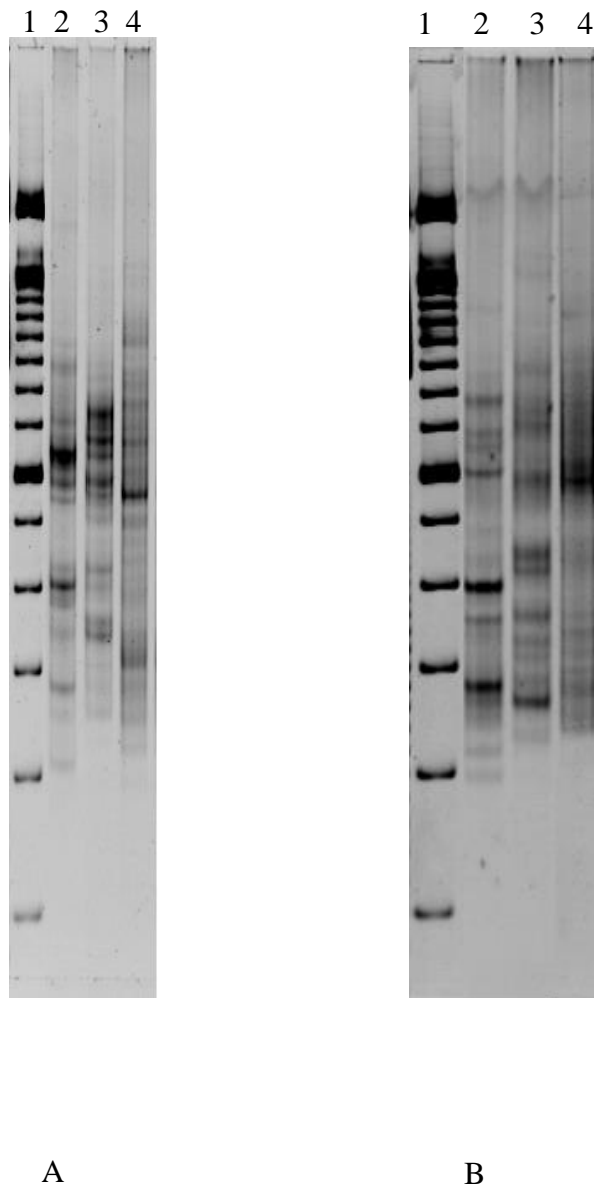
**Fig. 3:** PFGE of bacterial soil DNA extracted by indirect method from Montrond (A) and Châteaurenard (B) soil after enrichment. The ladder used was lambda ladder PFG marker.



fragment size, although this treatment was found to be efficient in preventing DNA damage in artificially FeIII-contaminated DNA solutions (results not shown). Thus, differences observed between soil samples treated with the same DNA extraction protocols (this study and Burgmann et al., 2001) do not appear to be due to specific soil properties that could lead to a physical, chemical or enzymatic degradation of the extracted bacterial DNA. These results were confirmed by results of the experiments in which *E. coli* cells were lysed according to the CsCl protocol used in this study in presence of Montrond soil supernatant (first centrifugation supernatant in the bacterial cell extraction protocol, see M&M section). The DNA smear indicating that these DNA fragments from *E. coli* ranged in size between 600 and 900 kb in the control (pure culture of *E. coli*) were not modified by extracted soil components present in the supernatant (results not shown). These results confirm that the extraction method and DNA size fragments were not affected by soil components. Nevertheless, the extraction protocol efficiency could vary according to bacterial diversity and / or physiological state. Bacteria of soil that live in a dormant state most of the time (Christensen et al., 1999) would contain DNA more sensitive to mechanical shearing than growing bacteria. Another hypothesis deals with absence of intact chromosomal DNA in dormant soil bacteria that might not spare the energy needed to maintain genome integrity under starvation conditions. Differences detected in the size of extracted DNA between soils might reflect the relative bacterial activity, which might be higher in Châteaurenard soil than in Montrond soil or presence of a different bacterial community in the 2 soils. However, a change in environmental conditions, such as nutrient input, would rapidly induce DNA repair systems to return the cell to a functioning state. This hypothesis was tested experimentally by adding nutrients, which could activate bacteria, to soil samples.

### **3.3 Enrichment of soil bacteria**

Activation of bacterial metabolism was carried out by adding a soil extract solution (Pochon and Tardieux, 1962), which provides the nutrients that are naturally present in the soil, to the soil samples. The use of a soil extract solution instead of other organic carbon sources was expected to limit the possible microbial selection bias. After incubation at 28°C for 60 h, the bacterial cells were extracted and DNA isolated and purified according to the cesium chloride density gradient protocol. With the Montrond soil, soil extract solution enrichment resulted in a qualitative increase of extracted DNA compared to the control soil. DNA fragments ranged in size between 25 and 400 kb with the maximum amount of DNA between 50 and 150 kb (**Fig. 3**). In this soil, the average size of the DNA fragments increased



**Fig. 4: Bacterial community structure of Montrond (A) and Châteaurenard (B) soils evaluated by RISA.** Lane 1: 100 bp DNA Ladder (Invitrogen), lane 2 : DNA extracted from enriched cells, lane 3: indirect DNA extraction, lane 4: direct DNA extraction.

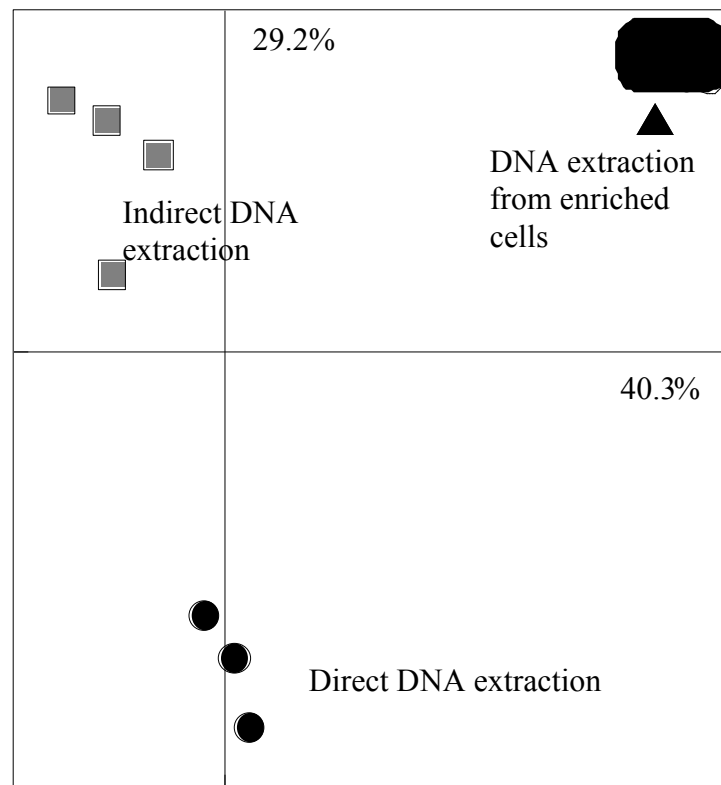
significantly with the enrichment treatment. These results suggest that conditions leading to either modification of bacterial community or an *in situ* activation of bacterial metabolism will lead to the extraction of longer DNA fragments. Quality of DNA recovered after an enrichment step with the Montrond soil sample was similar to that of Châteaurenard soil without enrichment. However, incubation of a soil sample from Châteaurenard with the soil extract solution did not lead to an increase of DNA extraction yield and the DNA smears on agarose gel after pulsed field gel electrophoresis did not indicate an increase in DNA fragment size.

The lack of effect on the Châteaurenard soil DNA size and the observation that the mean DNA length in soil extract augmented Montrond soil remained shorter than that of *in vitro* grown bacteria inoculated into soil supports the hypothesis that only a fraction of the bacterial community was stimulated by the soil extract solution. If this hypothesis is correct, then the bacterial community structure should be modified during the enrichment process.

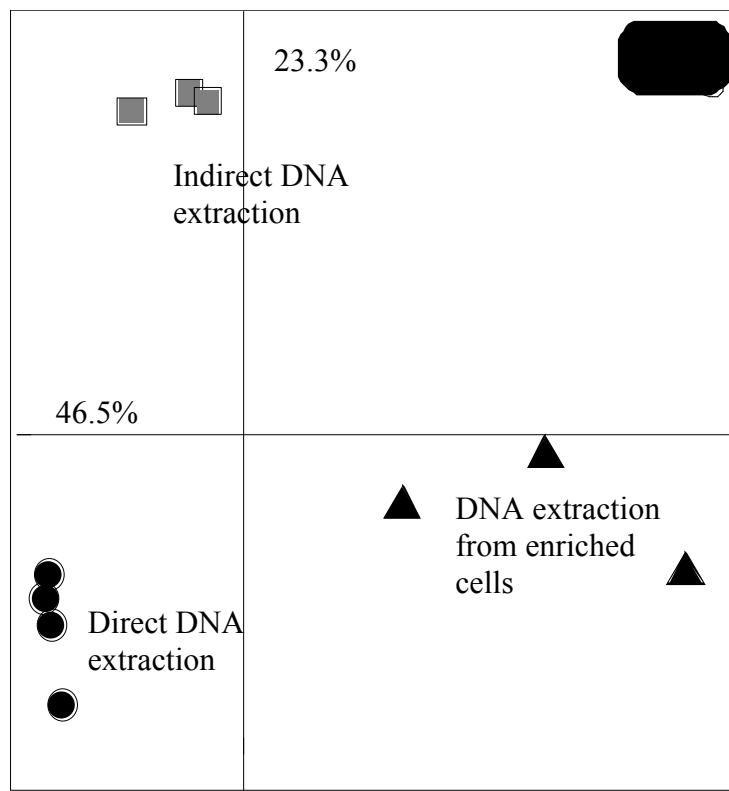
### **3.4 Structure of the soil bacterial community**

The level of the bias caused by the enrichment step on the bacterial community structure was studied by using the RISA approach with template DNA originating from enriched and control soil bacteria. DNA extracted from the same soil samples by the *in situ* lysis direct method was also tested. Results presented in **Fig. 4** indicated that for both soils, the three various conditions tested did not have similar RISA fingerprints. The direct method characterized by the *in situ* bacteria lysis treatment before DNA extraction produced RISA patterns significantly different from the samples where the bacterial extraction preceded the cell lysing. These differences could be partly due to presence of extracellular DNA co-extracted with DNA released from lysed bacteria during the direct extraction method. Fingerprint differences could also result from DNA desorption and lysis severity that would have different efficiencies depending on the protocol used (Martin-Laurent et al., 2002).

RISA fingerprints obtained with template DNA isolated from the enriched and non enriched soils via the indirect method were also different confirming that the incubation step modified the bacterial community structure. Interestingly, these differences between the two treatments were detected for both soils tested. The Châteaurenard soil for which quality of extracted DNA did not seem affected by the soil extract solution enrichment step exhibited a change in microbial composition. Statistical analysis of RISA fingerprints in terms of appearance-disappearance of DNA fragments can evaluate the importance of the various conditions (**Fig. 5**). Axis F1 in the PCA factorial map (**Fig. 5**) explained 40 and 46 % of the



A



B

**Fig. 5:** Statistical analysis of data profiles performed by principal components analysis (PCA) for Montrond (A) and Châteaurenard (B) soils.

variance in the data for Montrond and Châteaurenard soil DNA, respectively. These F1 axis allowed to differentiate enriched and unenriched soils. 23 and 29 % of the variance were explained on the F2 axis for Châteaurenard and Montrond soils, respectively distinguishing direct and indirect DNA extraction methods. Clearly, the RISA patterns were more affected by the soil enrichment treatment than by the use of the direct or indirect DNA extraction method. Previous studies based on PCR and hybridization with group specific oligonucleotide probes concluded similarly that soil community structure based on DNA extracted by direct or the indirect method varied little (Courtois et al., 2001). The RISA fingerprinting results presented here indicate that the enrichment treatment led to a significant change among the dominant bacteria of the microbial community for both soils tested. Conditions between in vitro and in situ growth result in differences in bacterial species discovered (Voget et al., 2003). Controlled enrichment could lead to the possibility of increasing the proportion of some bacterial taxonomic groups or of some bacteria with catabolic or secondary metabolite products. In this case, the enrichment approach combined with a cell bacterial extraction by the Nycodenz method and a cesium chloride gradient based DNA extraction and purification method might recover novel 300 kb long DNA fragments. The enrichment selection bias could, thus, be exploited to improve the detection and recovery of high value metagenomic DNA libraries (Entcheva et al., 2001 ; Knietsch et al., 2003).

#### **4 ACKNOWLEDGMENTS**

This work was financially supported by the European Community project METACONTROL (5th PCRT, Number: QLK3-CT-2002-02068) and by the Rhône-Alpes region (Thématiques Prioritaires 2003-2005, Sciences Analytiques Appliquées).

We thank P Robe and P Pujic (LibraGen) for their technical assistance, helpful discussions and comments on the manuscript.

#### **5 REFERENCES**

- Bakken, L.R., 1985. Separation and purification of bacteria from soil. Appl. Environ. Microbiol. 49, 1482-1487.
- Bakken, L.R. Olsen, R.A., 1989. DNA-content of soil bacteria of different cell size. Soil Biol. Biochem. 21, 789-793.
- Burgmann, H., Pesaro, M., Widmer, F. Zeyer, J., 2001. A strategy for optimizing quality and

- quantity of DNA extracted from soil. J. Microbiol. Methods. 45, 7-20.
- Christensen, H., Hansen, M. Sorensen, J., 1999. Counting and Size Classification of Active Soil Bacteria by Fluorescence In Situ Hybridization with an rRNA Oligonucleotide Probe. Appl. Environ. Microbiol. 65, 1753-1761.
- Courtois, S., Cappellano, C.M., Ball, M., Francou, F.X., Normand, P., Helynck, G., Martinez, A., Kolvek, S.J., Hopke, J., Osburne, M.S., August, P.R., Nalin, R., Guerineau, M., Jeannin, P., Simonet, P. Pernodet, J.L., 2003. Recombinant Environmental Libraries Provide Access to Microbial Diversity for Drug Discovery from Natural Products. Appl. Environ. Microbiol. 69, 49-55.
- Courtois, S., Frostegard, A., Goransson, P., Depret, G., Jeannin, P. Simonet, P., 2001. Quantification of bacterial subgroups in soil: comparison of DNA extracted directly from soil or from cells previously released by density gradient centrifugation. Environ. Microbiol. 3, 431-439.
- Daniel, R., 2004. The soil metagenome - a rich resource for the discovery of novel natural products. Curr Opin Biotechnol 15, 199-204.
- Doolittle, W.F., 1999. Phylogenetic classification and the universal tree. Science 284, 2124-2129.
- Entcheva, P., Liebl, W., Johann, A., Hartsch, T. Streit, W.R., 2001. Direct Cloning from Enrichment Cultures, a Reliable Strategy for Isolation of Complete Operons and Genes from Microbial Consortia. Appl. Environ. Microbiol. 67, 89-99.
- Gabor, E.M., de Vries, E.J. Janssen, D.B., 2003. Efficient recovery of environmental DNA for expression cloning by indirect extraction methods. FEMS Microbiol. Ecol. 44, 153-163.
- Joseph, S.J., Hugenholtz, P., Sangwan, P., Osborne, C.A. Janssen, P.H., 2003. Laboratory Cultivation of Widespread and Previously Uncultured Soil Bacteria. Appl. Environ. Microbiol. 69, 7210-7215.
- Knietsch, A., Waschowitz, T., Bowien, S., Henne, A. Daniel, R., 2003. Construction and Screening of Metagenomic Libraries Derived from Enrichment Cultures: Generation of a Gene Bank for Genes Conferring Alcohol Oxidoreductase Activity on *Escherichia coli*. Appl. Environ. Microbiol. 69, 1408-1416.
- MacNeil, I.A., Tiong, C.L., Minor, C., August, P.R., Grossman, T.H., Loiacono, K.A., Lynch, B.A., Phillips, T., Narula, S., Sundaramoorthi, R., Tyler, A., Aldredge, T., Long, H., Gilman, M., Holt, D. Osburne, M.S., 2001. Expression and isolation of antimicrobial small molecules from soil DNA libraries. J Mol Microbiol Biotechnol 3, 301-308.

- Martin-Laurent, F., Philippot, L., Hallet, S., Chaussod, R., Germon, J.C., Soulas, G. Catroux, G., 2001. DNA extraction from soils: old bias for new microbial diversity analysis methods. *Appl. Environ. Microbiol.* 67, 2354-239.
- Netto, L.E., Ferreira, A.M. Augusto, O., 1991. Iron(III) binding in DNA solutions: complex formation and catalytic activity in the oxidation of hydrazine derivatives. *Chem. Biol. Interact.* 79, 1-14.
- Normand, P., Ponsonnet, C., Nesme, X., Neyra, M. Simonet, P., 1996. ITS analysis of prokaryotes. In : Akkermans, A.D.L., Van Elsas, F.J. de Bruijn, F.J. (Eds.), *Molecular Microbial Ecology Manual*. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 1-12.
- Ogram, A., Sayler, G.S. Barkay, T., 1987. The extraction and purification of microbial DNA from sediments. *J. Microbiol. Methods.* 7, 57-66.
- Picard, C., Ponsonnet, C., Paget, E., Nesme, X. Simonet, P., 1992. Detection and enumeration of bacteria in soil by direct DNA extraction and polymerase chain reaction. *Appl. Environ. Microbiol.* 58, 2717-2722.
- Pochon, J. Tardieux, P., 1962. . In : *Techniques d'analyse en microbiologie du sol*. Edition de la Tourelle, Saint Mandé, 111p.
- Poly, F., Gros, R., Jocteur-Monrozier, L. Perrodin, Y., 2002. Short-term changes in bacterial community fingerprints and potential activities in an alfisol supplemented with solid waste leachates. *Environ. Sci. Technol.* 36, 4729-434.
- Richaume, A., Steinberg, C., Jocteur Monrozier, L. Faurie, G., 1993. Difference between direct and indirect enumeration of soil bacteria: influence of soil structure and cell location. *Soil Biol. Biochem.* 25, 641-643.
- Robe, P., Nalin, R., Capellano, C., Vogel, T.M. Simonet, P., 2003. Extraction of DNA from soil. *Eur. J. Soil. Biol.* OCT-DEC 2003, 183-190.
- Rossello-Mora, R. Amann, R., 2001. The species concept for prokaryotes. *Fems. Microbiol. Rev.* 25, 39-67.
- Schwinghamer, E.A., 1980. A method for improved lysis of some Gram-negative bacteria. *FEMS Microbiol. Lett.* 7, 157-162.
- Seow, K.T., Meurer, G., Gerlitz, M., Wendt-Pienkowski, E., Hutchinson, C.R. Davies, J., 1997. A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: a means to access and use genes from uncultured microorganisms. *J. Bacteriol.* 179, 7360-7368.
- Thioulouse, J., Chessel, D., Doledec, S. Olivier, J.M., 1997. ADE-4: a multivariate analysis

- and graphical display software. *Stat. Comput* 7, 75-83.
- Torsvik, V., Daae, F.L., Sandaa, R.A. Ovreas, L., 1998. Novel techniques for analysing microbial diversity in natural and perturbed environments. *J. Biotechnol.* 64, 53-62.
- Torsvik, V. Ovreas, L., 2002. Microbial diversity and function in soil: from genes to ecosystems. *Curr. Opin. Microbiol.* 5, 240-245.
- Voget, S., Leggewie, C., Uesbeck, A., Raasch, C., Jaeger, K.E. Streit, W.R., 2003. Prospecting for Novel Biocatalysts in a Soil Metagenome. *Appl. Environ. Microbiol.* 69, 6235-6242.
- Wang, G.Y., Graziani, E., Waters, B., Pan, W., Li, X., McDermott, J., Meurer, G., Saxena, G., Andersen, R.J. Davies, J., 2000. Novel natural products from soil DNA libraries in a streptomycete host. *Org. Lett.* 2, 2401-2404.
- Zhang, H., Chen, H.T. Glisin, V., 2003. Isolation of DNA-free RNA, DNA, and proteins by cesium trifluoroacetate centrifugation. *Biochem. Biophys. Res. Commun.* 312, 131-137.
- Zolg, J.W., Scott, E.D. Wendlinger, M., 1988. High salt lysates: a simple method to store blood samples without refrigeration for subsequent use with DNA probes. *Am. J. Trop. Med. Hyg.* 39, 33-40.



## CONCLUSIONS ET PERSPECTIVES

Il est classiquement rappelé dans la littérature que les traitements d'extraction de l'ADN à partir du sol sont responsables de la dégradation de l'ADN produisant des fragments dont la taille est généralement inférieure à 200 kb, très loin de celle obtenue à partir de cellules cultivées *in vitro*. Les objectifs dans ce chapitre visaient à déterminer si ces résultats étaient liés à un effet délétère de nature physique, chimique ou enzymatique de la matrice tellurique sur l'ADN et s'il était possible, le cas échéant de développer des protocoles permettant de s'affranchir de ces biais de façon à pouvoir cloner ultérieurement des fragments de très grande taille.

Le résultat le plus important de ces travaux a été de montrer que les techniques d'extraction ne sont pas les seuls facteurs responsables de la faible taille des fragments d'ADN extraits à partir des bactéries du sol. Les données établissent clairement que la qualité de l'ADN extrait est dépendante de l'état dans lequel celui-ci se trouve à l'intérieur des cellules microbiennes dans le sol. L'état physiologique des cellules est notamment responsable des différences notées entre deux sols. Celui le plus riche en matière organique (Châteaurenard), donc le plus favorable au développement bactérien, produit un ADN de meilleure qualité qu'un sol plus pauvre en matière organique (Montrond), dans lequel seul l'ajout de nutriments permet d'améliorer l'état de l'ADN extrait.

Les bactéries survivent la majeure partie du temps dans le sol dans un état de vie très ralenti (Christensen *et al.*, 1999; De Fede & Sexstone, 2001; Maraha *et al.*, 2004). Dans cet état de dormance, les organismes ne maintiendraient pas l'intégrité de leur ADN, économisant ainsi leur énergie. L'ajout de nutriments peut être perçu comme un stress suite à une modification environnementale, entraînant ainsi l'activation du système SOS de réparation d'ADN ou comme une stimulation métabolique permettant la réparation de l'ADN. La très importante capacité des bactéries à réparer l'ADN dégradé, notamment montré chez la bactérie *Deinococcus radiodurans*, conforte ces hypothèses. Les mécanismes de réparation de cette bactérie, similaire à ceux des autres procaryotes (dont ceux du sol), lui permettent de résister à des niveaux de radiation ionisante très élevés (Daly *et al.*, 1994). L'ADN fortement dégradé, présent dans un sol où les conditions sont peu favorables au développement bactérien (Montrond), est potentiellement réparable par les bactéries comme l'ont montré les

stimulations nutritives. L'absence d'amélioration de qualité de l'ADN extrait du sol de Châteaurenard suite à l'ajout de nutriments est quant à lui expliqué par un éveil métabolique continu dans ce sol, du à la présence de matière organique en quantité suffisante. Une majorité de bactéries dans ce sol ne serait donc pas en phase de dormance, ce qui expliquerait le maintien de l'intégrité de l'ADN chez ces organismes sans possibilité d'augmenter la taille des fragments extraits par stimulation nutritionnelle.

Dans l'un et l'autre des cas cette stimulation s'accompagne d'une modification significative de la communauté en favorisant certains taxons au détriment d'autres, biaisant ainsi l'analyse de la structure bactérienne initiale. Ces biais pourraient toutefois être mis à profit pour favoriser l'isolement de gènes spécifiques liés à la croissance de ces taxons particuliers (Hery *et al.*, 2005a; Smith *et al.*, 2001). La probabilité d'obtenir au moins un crible positif est dépendante de la taille de la banque construite, de la taille moyenne de l'insert métagénomique cloné, de la taille moyenne du (des) gène(s) recherché(s), de la taille des génomes et de l'abondance des espèces (Gabor *et al.*, 2003). Plusieurs milliers de clones doivent donc être criblés avant de détecter l'activité recherchée (Rondon *et al.*, 2000; Courtois *et al.*, 2003; Riesenfeld *et al.*, 2004a). Les méthodes d'enrichissement optimisent les chances de détecter des clones possédant les gènes ou l'activité recherchée.

Certains substrats spécifiques sont ainsi apportés aux cellules environnementales pour enrichir celles exprimant une activité spécifique avant la construction des banques métagénomiques (Entcheva *et al.*, 2001; Gabor *et al.*, 2003; Knietsch *et al.*, 2003). Contrairement aux enrichissements classiques effectués pour l'isolement d'une bactérie, les enrichissements sont dans ce cas effectués pour un mélange de cellules environnementales, conservant ainsi les consortiums bactériens pouvant être nécessaires (métabiosis) à la croissance de certaines bactéries (Schink, 2002). Malgré les biais introduits par les conditions de culture, l'enrichissement a ainsi permis l'isolement de gènes nouveaux (Entcheva *et al.*, 2001; Gabor *et al.*, 2003; Knietsch *et al.*, 2003). Une approche dérivée de ces enrichissements est l'extraction de l'information génétique uniquement des cellules bactériennes actives grâce au marquage isotopique au carbone  $^{13}\text{C}$  du substrat apporté. L'incorporation du carbone lourd  $^{13}\text{C}$  dans l'ADN des cellules métaboliquement actives (ADN lourd) permet sa séparation de l'ADN des cellules non actives (ADN léger) sur gradient de chlorure de césium. Le clonage d'ADN métagénomique purifié selon cette approche SIP (*stable isotope probing*) permet la construction de banques métagénomiques ciblées sur les populations actives métabolisant le substrat marqué (Dumont & Murrell, 2005; Dumont *et al.*, 2006; Wellington *et al.*, 2003). Une autre stratégie de criblage vise à l'enrichissement des informations métagénomiques

après construction de la banque. Uchiyama *et al.* (2005) ont ainsi développé un système de détection et isolement des clones comportant dans leur insert des promoteurs inductibles à la présence de composés aromatiques et les gènes intervenant dans leur catabolisme (Uchiyama *et al.*, 2005).

Aux vues de la diversité bactérienne présente dans l'environnement et de la taille des banques nécessaires à la couverture de cette diversité, des méthodes facilitant le criblage d'événements rares dans les banques métagénomique est ainsi devenu un besoin réel. Les méthodes « d'enrichissement » basées non plus sur l'expression d'une activité, mais sur la similarité des séquences manquent actuellement. Bien que la PCR soit une approche basée sur une similarité de séquence, cette dernière biaise l'accès à l'information génétique de la microflore tellurique en amplifiant les séquences dominantes. Aussi toutes les approches dérivant de la PCR sont sujettes aux mêmes biais. Par ailleurs, la PCR est généralement effectuée à partir d'un ADN extrait de l'environnement. Comme nous l'avons précédemment mentionné, les méthodes d'extraction et purification de l'ADN ainsi que le clonage et la transformation dans la cellule hôte entraîne des pertes d'informations génétiques.

**Une approche basée sur la similarité de séquence et évitant les biais liés à l'extraction et purification de l'ADN extrait de l'échantillon étudié, ainsi qu'à l'amplification PCR, donnerait accès à un nouveau degré d'information génétique.**

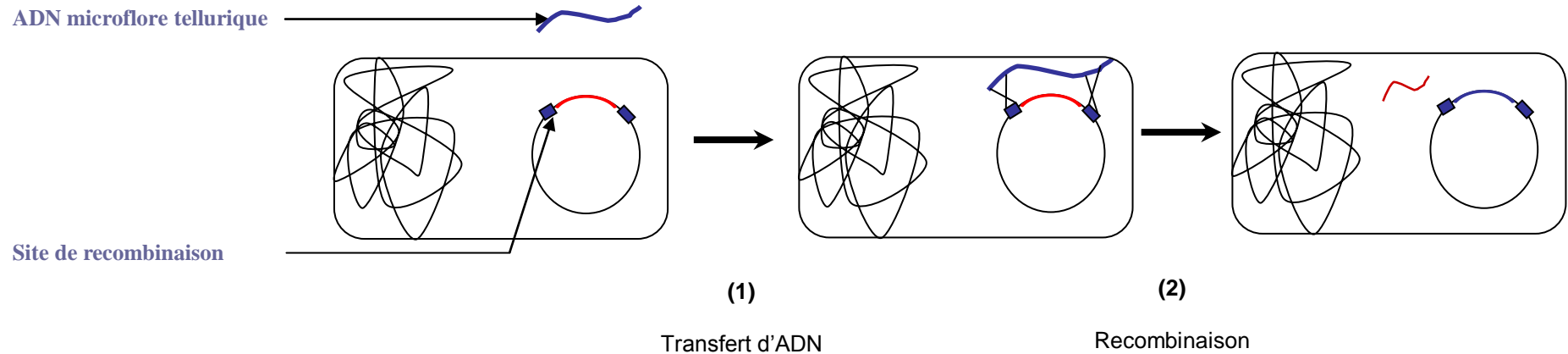


### **CHAPITRE III : RESULTATS**

#### **DEVELOPPEMENT D'UNE NOUVELLE APPROCHE METAGENOMIQUE : CAPTURE *IN SITU* DE L'ADN METAGENOMIQUE**

Dans la synthèse bibliographique (chapitre I) était souligné l'impact des méthodes d'extraction sur la pureté de l'ADN extrait, la taille des fragments d'ADN, ainsi que sur la représentativité de l'information génétique recueillie. Les biais liés à l'extraction et au clonage de l'ADN métagénomique ont notamment des répercussions sur la diversité des espèces présentes dans les banques métagénomiques construites (Luna *et al.*, 2006). Afin de compléter les informations génétiques ainsi collectées, nous proposons une voie alternative pour appréhender, étudier et exploiter le réservoir inexploré des ressources génétiques bactériennes. Dans cette nouvelle approche, une souche hôte (réceptrice) spécialement préparée génétiquement est directement introduite dans le sol pour piéger l'ADN de la microflore tellurique. L'extraction *ex situ* de l'ADN des bactéries est ainsi contournée, ouvrant un accès tout au moins théorique vers les 30 à 50 % d'informations génétiques jusqu'alors soit retenues par la matrice sol, soit perdues lors des processus de purification.

L'approche proposée est un processus en deux étapes, au cours duquel la souche réceptrice internalise l'ADN dans son cytoplasme avant de l'intégrer dans son génome (**Figure 1**), permettant ainsi la stabilisation et la réplication de l'ADN provenant de la microflore tellurique. Cette intégration a lieu par double recombinaison (double *crossing over*) entre l'ADN métagénomique et des séquences nucléotidiques (sites de recombinaison) introduites dans la souche réceptrice, permettant ainsi le piégeage d'une information génétique ciblée. De même que pour les approches PCR, l'accès à la diversité des informations génétiques sera permise par la définition de deux sites de recombinaison dans les zones conservées des gènes, opérons ou voies de biosynthèses ciblées. Cette stratégie de capture basée sur la similarité de séquences permettra l'accès à une diversité ciblée en une seule étape, contrairement aux banques métagénomiques classiques pour lesquelles les séquences d'ADN de la microflore tellurique sont d'abord collectées avant d'être criblées pour identifier les gènes d'intérêt.



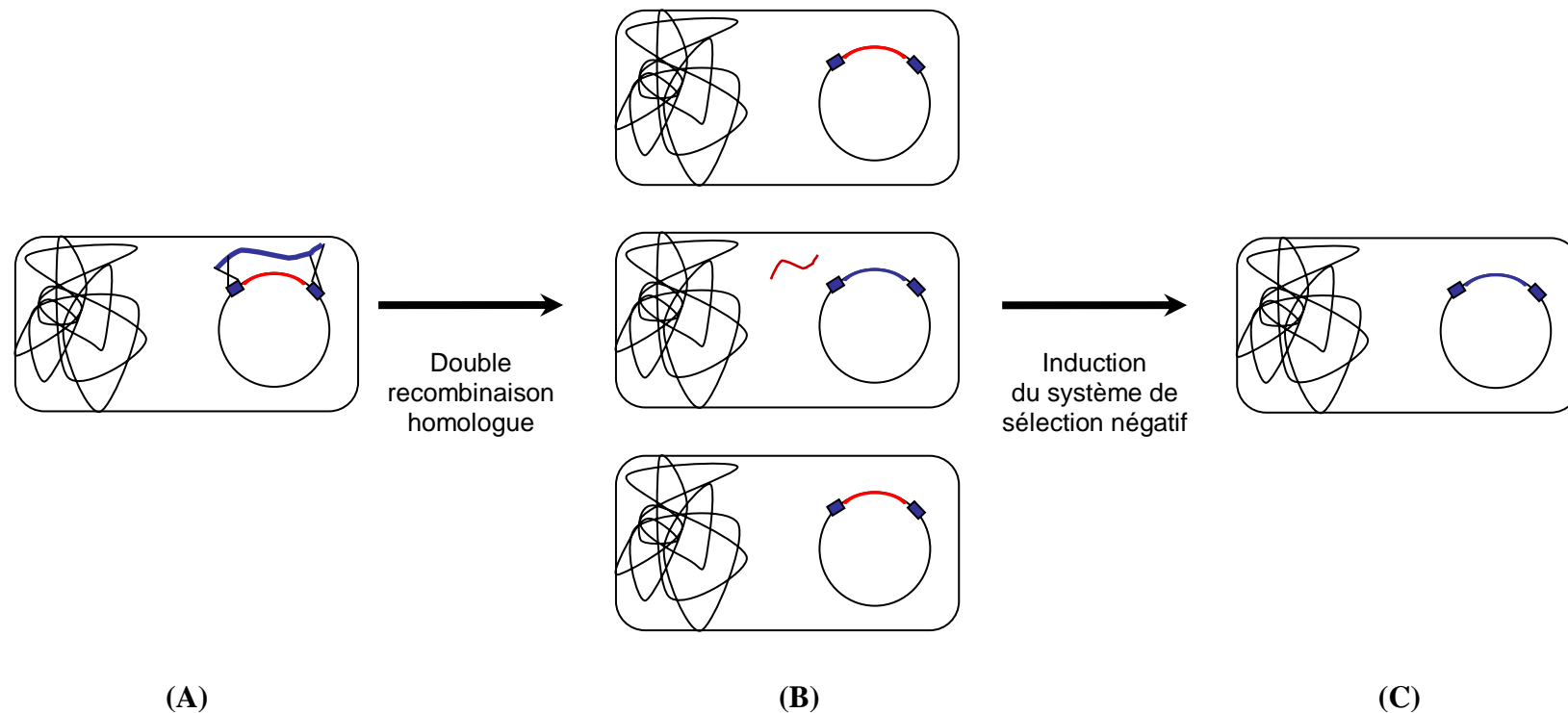
**Figure 1 : Mécanisme de piégeage de l'ADN de la microflore tellurique dans une souche réceptrice**

La capture de l'ADN de la microflore tellurique dans une souche réceptrice est un processus en deux étapes : **(1)** l'ADN est internalisé dans le cytoplasme de la cellule réceptrice par un mécanisme de transfert d'ADN **(2)** l'ADN entrant est intégré dans le génome de la souche réceptrice (ici un plasmide) grâce à des sites de recombinaison présents dans la souche réceptrice. Si une homologie de séquence est suffisante entre l'ADN entrant et les sites de recombinaison, la double recombinaison (double *crossing over*) permettra l'intégration du fragment ADN métagénomique.

La faisabilité d'une telle approche est appuyée par l'existence dans l'environnement de transferts naturels d'information génétique entre espèces bactériennes différentes. Plusieurs processus sont responsables des échanges génétiques entre bactéries : la transduction, la conjugaison, la transformation naturelle (Bertolla & Simonet, 1999; Davison, 1999) ainsi que l'électrotransformation naturelle. Cette variante de la transformation naturelle permet l'entrée de l'ADN à l'intérieur de la cellule suite à un choc électrique tel que la foudre. Cette voie d'acquisition de l'ADN a pour la première fois été mise en évidence dans l'environnement par Demanèche *et al.* (2001) à des fréquences atteignant  $10^{-4}$ - $10^{-5}$  dans le sol (Ceremonie *et al.*, 2004)

Par ailleurs, les cellules bactériennes possèdent des systèmes permettant l'intégration de l'ADN par recombinaison dans leurs génomes. Le plus étudié est la voie de recombinaison RecBCD de *Escherichia coli*. Ce système intervient dans la réponse SOS pour la réparation par recombinaison de l'ADN endommagé (Michel, 2005). L'ADN simple brin ou double brin linéaire entrant dans la cellule est pris en charge par un ensemble de protéines, puis intégré dans le génome (plasmidique ou chromosomique) lorsqu'une homologie de séquences est suffisante pour amorcer la recombinaison avec le génome récepteur (Matic *et al.*, 1996). Ces événements de recombinaison ont lieu pour des régions de forte homologie, de 25 à 200 pb suivant les espèces (Thomas & Nielsen, 2005) et sont beaucoup plus élevées lorsque deux régions d'homologie sont présentes au lieu d'une seule (de Vries & Wackernagel, 2002).

Cette approche de capture de gènes nécessite cependant un système permettant l'identification des cellules réceptrices ayant intégré dans leur génome l'information génétique ciblée. La sélection est un moyen pour séparer les cellules hôtes ayant acquis un fragment d'ADN des autres cellules, par la croissance des seules cellules ayant acquis le fragment d'ADN (sélection positive), ou au contraire par la mort des cellules ayant intégré le fragment d'ADN (sélection négative) (Miki & McHugh, 2004). Ces deux types de sélection, positive et négative, peuvent être conditionnelles, la croissance ou l'absence de croissance de la souche n'est alors entraînée que par l'ajout d'un substrat (Miki & McHugh, 2004). Les gènes de résistance aux antibiotiques permettent par exemple une sélection positive, alors que les gènes conférant une auxotrophie pour un substrat ou les gènes codant une fonction toxique pour la cellule (gènes toxiques) sont classés dans la sélection négative.



**Figure 2 : Principe de la contre sélection pour l'isolement des souches réceptrices transformées / recombinées**

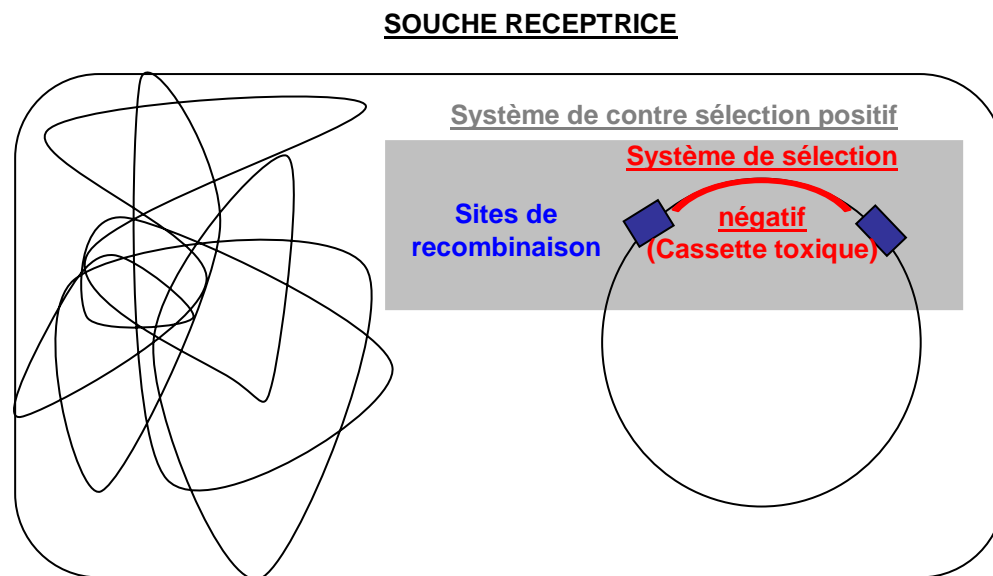
(A) La double recombinaison homologue a lieu uniquement si une homologie de séquence suffisante est présente entre l'ADN entrant (bleu) et les sites de recombinaison (bleu) clonés dans la souche réceptrice. (B) Pour différencier les cellules réceptrices ayant intégré l'ADN entrant des autres cellules réceptrices, un système de sélection négatif inductible (en rouge), constitué de gènes exprimant une activité toxique pour la cellule en présence d'inducteur, a été ajouté entre les sites de recombinaison. La double recombinaison homologue conduit à l'élimination du système de sélection négatif (ou cassette toxique). (C) Suite à l'apport du substrat (inducteur) permettant la cassette toxique, seule les souches ayant délété la cassette toxique survivront. Ainsi, seules les souches transformées/ recombinées survivent.



Ces deux approches sélectives ne sont cependant exploitables que si le phénotype apporté par le fragment d'ADN cloné est connu, ou si le fragment d'ADN peut être modifié par l'ajout d'un gène de sélection. Dans notre approche où des fragments d'ADN métagénomique inconnus sont clonés, un système de contre sélection positive a donc été adopté: la souche hôte porte un système de sélection négative conditionnelle (ou inductible), qui provoque sa mort suite à l'apport d'un substrat (inducteur). L'intégration du fragment d'ADN ciblé par double recombinaison conduit à l'élimination du système de sélection négative. Ainsi seules les souches recombinées survivront à l'apport de l'inducteur (**Figure 2**). Cette méthode permet ainsi la sélection des souches réceptrices ayant récupéré le fragment d'ADN, sans connaissance préalable ou modification du fragment d'ADN intégré.

Des systèmes de contre sélection positive ont déjà été développés pour l'inactivation de gènes, via l'échange allélique entre un gène fonctionnel et un fragment d'ADN portant le même gène muté. Le remplacement du gène fonctionnel par le gène muté est sélectionné d'une part grâce à un gène de résistance antibiotique apporté par le gène muté, d'autre part grâce à un gène toxique qui doit être éliminé lors d'une double recombinaison entre gène fonctionnel et gène muté (Parish & Stoker, 2000; Zhang *et al.*, 2006). Ces systèmes, développés pour l'étude de la fonction des gènes mutés, sont cependant employés pour des conditions optimales de recombinaison. Les zones présentent une totale similarité entre ADN donneur et ADN receveur. De plus, les événements de transformation / recombinaison peuvent atteindre en routine des fréquences de  $10^{-3}$ - $10^{-4}$  *in vitro* grâce à des systèmes additionnels favorisant la recombinaison (Murphy, 1998).

Les conditions d'utilisation de notre outil engendrent des fréquences de transformation et recombinaison beaucoup plus faibles du fait d'une plus faible efficacité de la transformation *in situ*, de la dilution des gènes cibles dans le métagénome, et des divergences de séquences des cibles entre ADN donneur et régions clonées dans la souche réceptrice. Ce nouvel outil métagénomique implique donc des événements de transformation / recombinaison avec l'ADN métagénomique à des fréquences faibles, voire même très faibles (inférieures à  $10^{-4}$ ) justifiant le développement de systèmes de contre sélection des cellules non transformées / recombinées les plus efficaces possibles, de façon à détecter les événements de capture les plus rares.



**(C) (sous chapitre III-4)**

**(B) (sous chapitre III-2)**

**(A) (sous chapitre III-1 et III-3)**

**Figure 3 : Nomenclature et points abordés pour la conception de la souche réceptrice**

La nouvelle approche métagénomique basée sur le piégeage de gène de la microflore tellurique repose sur la conception d'une souche réceptrice modifiée pour l'accomplissement de cette capture : la souche réceptrice doit être apte à l'intégration et la recombinaison d'ADN étranger, et possède un système de contre sélection positif porté par un plasmide, composé deux sites de recombinaison et d'un système de sélection négatif, également nommé « cassette toxique ». Cette cassette toxique est constituée de gènes codant des fonctions toxiques pour la cellule, placés sous le contrôle d'un système de régulation inductible.

Chacun des points développés pour la conception de la souche réceptrice sont soulignés :

- (A) le système de contrôle de l'expression des gènes toxiques a d'abord été étudié
- (B) le système de contre sélection positif développé a été évalué pour son aptitude au piégeage de l'ADN de la microflore tellurique
- (C) les espèces bactériennes exploitables comme souche réceptrices ont été recherchées et évaluées

La nouvelle approche métagénomique proposée dans cette thèse repose sur la conception d'une souche réceptrice capable de piéger l'ADN de la microflore tellurique. Cette souche réceptrice doit être capable d'internaliser et intégrer l'ADN étranger à une fréquence élevée, et doit posséder un système permettant l'isolement des seules souches réceptrices ayant incorporé cet ADN étranger. Dans ce chapitre, plusieurs points ont donc été abordés pour la conception de ce nouvel outil métagénomique (**Figure 3**). Un système de contre sélection efficace a été développé, afin que le seuil de détection des souches réceptrices ayant capturé l'ADN métagénomique ciblé (souches réceptrices recombinées) soit le plus bas possible. Le développement du système de contre sélection a notamment nécessité l'étude de systèmes d'expression inductible pour le contrôle de l'expression des gènes toxiques qui composent le système, afin d'optimiser l'efficacité tueuse de la cassette toxique finale (sous chapitre III-1). Les premiers essais expérimentaux (sous chapitre III-2) et les optimisations complémentaires nécessaires pour la validation de l'approche sont également présentés (sous chapitre III-3). Dans un dernier sous chapitre expérimental (sous chapitre III-4), des essais pour élargir le champ des souches réceptrices potentielles pour leur application dans cette nouvelle voie d'accès à l'ADN métagénomique ont également été réalisés et sont présentés.

## **SOUS CHAPITRE III-1 :**

### **ETUDE ET DEVELOPPEMENT DE SYSTEMES DE REGULATION POUR LE CONTROLE DE L'EXPRESSION DE GENES TOXIQUES**

#### **1 INTRODUCTION**

Les gènes toxiques sont par définition des gènes dont l'expression entraîne des perturbations majeures de la cellule hôte jusqu'à sa mort. Ces gènes ont fait l'objet d'exploitations biotechnologiques variées ces dernières années : comme outil de contre sélection dans le cadre de remplacement allélique (Zhang *et al.*, 2006), pour éviter la dissémination dans la microflore environnante de l'information génétique de bactéries Génétiquement Modifiées (GM) libérées volontairement ou accidentellement dans l'environnement (Ahrenholtz *et al.*, 1994; Knudsen & Karlstrom, 1991; Torres *et al.*, 2000), et dans l'étude des transferts horizontaux de gènes pour contre sélectionner les souches donatrices (Sengelov & Sorensen, 1998). Dans le contexte de biorémédiation bactérienne, les gènes toxiques ont également été envisagés pour entraîner la mort des cellules bactériennes introduites dans l'environnement une fois leurs fonctions de dégradation accomplies (Jensen *et al.*, 1993; Ronchel *et al.*, 1995; Ronchel & Ramos, 2001).

Ces gènes aux propriétés toxiques ont des origines et des fonctions variées. Certains sont originaires de bactériophages, comme le gène E exprimé par le phage  $\phi$ X174 lors de son cycle lytique (van Dillewijn *et al.*, 2004). Beaucoup sont naturellement présents chez les bactéries. Les bactériocines par exemple exprimées naturellement dans le contexte de compétition entre bactéries plus ou moins proches phylogénétiquement (Ellermeier *et al.*, 2006; Riley & Wertz, 2002) montrent des propriétés bactéricides exploitables à des fins technologiques. Les gènes exprimant les colicines, parmi les plus étudiées des bactériocines, constituent notamment des gènes facilement manipulables, un seul gène étant nécessaire à l'expression d'une activité toxique, contrairement à d'autres bactériocines pour lesquelles la maturation de la protéine toxique nécessite l'expression de plusieurs gènes (Riley & Wertz, 2002). D'autres gènes impliqués dans la mort programmée des cellules (Engelberg-Kulka *et*

*al.*, 2005), dans les systèmes de maintenance de plasmides (Afif *et al.*, 2001; Molin *et al.*, 1993), ainsi que tous les systèmes nucléasiques tels que les systèmes de restriction (Torres *et al.*, 2000; Torres *et al.*, 2003) ou encore les systèmes de modification de l'ADN (Setlow *et al.*, 1992), ont également fait l'objet d'une reconversion pour des applications technologiques nécessitant des propriétés létales.

L'efficacité tueuse des gènes toxiques est cependant très variable d'un gène à l'autre. Pour une construction génétique dans une souche bactérienne donnée, une partie de la population est capable de survivre en dépit de l'expression toxique. Ce taux d'échappement peut varier de  $10^{-2}$  à  $10^{-5}$  lorsque le gène toxique est cloné sur un plasmide dans un système de régulation simple (Ahrenholtz *et al.*, 1994; Aizenman *et al.*, 1996; Knudsen & Karlstrom, 1991; Setlow *et al.*, 1992). Ces bactéries échappent à l'action toxique suivant différents mécanismes dont la mutation du gène (Diaz *et al.*, 1994; Recorbet *et al.*, 1993; Torres *et al.*, 2000) ou du système contrôlant son expression (Knudsen *et al.*, 1995; Knudsen & Karlstrom, 1991; Torres *et al.*, 2003). Afin de diminuer cette fréquence d'échappement, des stratégies ont été développées consistant à dupliquer le gène toxique (Jensen *et al.*, 1993; Knudsen *et al.*, 1995; Knudsen & Karlstrom, 1991) ou à cloner deux gènes toxiques différents (Torres *et al.*, 2003). Cependant l'efficacité des gènes toxiques à tuer la cellule hôte dépend également du système de régulation de son expression, notamment de la précision du contrôle de l'expression exercée par l'expérimentateur. Lors du clonage de gènes toxiques dans un système d'expression inductible, la persistance d'une expression résiduelle en absence d'inducteur a été relevée (Lanzer & Bujard, 1988). Une diminution de cette expression basale du gène toxique permet de multiplier jusqu'à 100 fois l'efficacité tueuse du gène toxique (Szafranski *et al.*, 1997). La construction d'une cassette toxique inductible requiert donc un contrôle optimal de l'expression toxique.

Les systèmes de contrôle de l'expression par régulation négative, moins sujets aux mutations (Knudsen & Karlstrom, 1991), reposent sur l'inhibition continue de la transcription. Des protéines, exprimées constitutivement par des gènes répresseurs, se fixent sur des séquences nucléotidiques spécifiques, dites opératrices, du promoteur. Leur fixation entraîne l'inhibition de la transcription des gènes situés en aval du promoteur. L'ajout d'un inducteur permet la dérépression et / ou l'activation de la transcription (Griffiths *et al.*, 1999). Plusieurs systèmes de contrôle à régulation négative ont été développés pour leur utilisation en biologie moléculaire, dont les systèmes dérivés de l'opéron lactose et de l'opéron arabinose. Le système de contrôle dérivé de l'opéron lactose a été très largement étudié. Des modifications de ce système ont été réalisées afin d'optimiser d'une part la quantité de protéines produites et

d'autre part de diminuer les expressions résiduelles en absence d'inducteur (Lanzer & Bujard, 1988). Ce dernier est constitué d'un promoteur de type  $P_{lac}$ , dont l'activité est contrôlée par le gène répresseur *lacI*. L'expression des gènes clonés sous  $P_{lac}$  est permise par dérépression en présence d'une molécule inductrice, l'IPTG (isopropyl  $\beta$ -D-thiogalactopyranoside) (Griffiths *et al.*, 1999). Le système de contrôle dérivé de l'opéron arabinose est constitué d'un promoteur  $P_{bad}$  dont l'activité est régulée par le gène répresseur *araC*. Une particularité de ce type de système est la répression / activation entraînée par la protéine AraC en fonction de l'absence ou de la présence de l'inducteur arabinose. Ce qui permet au système d'une part une très faible expression basale en absence d'inducteur, d'autre part une forte expression en condition d'induction (Guzman *et al.*, 1995). Les promoteurs jouent un rôle important dans le contrôle de l'expression, d'une part dans l'efficacité de la répression par l'emplacement des séquences opératrices, d'autre part dans l'efficacité de la transcription par leur affinité avec l'ARN polymérase (Lanzer & Bujard, 1988).

Dans cette partie du projet, notre objectif a été de développer une cassette constituée de gène(s) toxique(s) dans un système d'expression inductible. Cette cassette toxique, qui fera partie intégrante d'un système de contre sélection positive des cellules bactériennes, doit être la plus performante possible afin de détecter des événements de transformation / recombinaison intervenant à très basse fréquence. Cet objectif a été abordé en plusieurs étapes, la première consistant à optimiser le contrôle de l'expression de gènes toxiques. Pour ce faire deux gènes toxiques ont été clonés sous différents systèmes de contrôle à régulation négative, afin de sélectionner celui permettant le meilleur compromis entre une expression résiduelle minimale en absence d'inducteur et une mortalité des cellules hôtes maximale en sa présence. Plusieurs promoteurs dérivés des systèmes de régulation lactose et arabinose ont ainsi été testés pour leur aptitude au contrôle de l'expression de gènes toxiques.

Deux promoteurs dérivés de  $P_{lac}$  ont été testés. Le premier est le promoteur hybride  $P_{tac/lacUV5}$  décrit par Moralès *et al.* (1991) comme étant finement régulé. Ce dernier a déjà été utilisé pour cloner des gènes ayant un effet néfaste pour la cellule lorsqu'ils sont exprimés (Morales *et al.*, 1991). Le deuxième est le promoteur synthétique  $P_{A1-O3/O4}$  construit par Bujard *et al.* et décrit ultérieurement dans la littérature comme le promoteur  $P_{A1-lacO1}$  (Lutz & Bujard, 1997). Grâce à sa très faible expression basale en absence d'inducteur et son expression élevée en sa présence, ce promoteur a été très utilisé pour cloner le gène marqueur *gfp* (Andersen *et al.*, 1998; Leveau & Lindow, 2001), ainsi que de nombreux gènes toxiques (Jensen *et al.*, 1993; Knudsen *et al.*, 1995; Ronchel & Ramos, 2001; van Dillewijn *et al.*,

**Tableau 1 : Liste des souches *E. coli* utilisées et construites**

Souche <i>E.coli</i>	Génotype	Résistance	Source
<b>DH5</b>	F <sup>-</sup> $\phi$ 80lacZ $\Delta$ M15 $\Delta$ (lac-argF)U169 deoR recA1 endA1 hsdR15(rk <sup>-</sup> , mk <sup>+</sup> ) phoA supE44 thi-1 gyrA96 relA1 $\lambda^-$		Invitrogen
<b>DH5Z1</b>	F <sup>-</sup> $\phi$ 80lacZ $\Delta$ M15 $\Delta$ (lac-argF)U169 deoR recA1 endA1 hsdR15(rk <sup>-</sup> , mk <sup>+</sup> ) phoA supE44 thi-1 gyrA96 relA1 $\lambda^-$ spr lacI <sup>q</sup> tetR	Sp	Lutz <i>et al.</i> , 1997
<b>EC100™</b>	F <sup>-</sup> $\phi$ 80dlacZ $\Delta$ M15 $\Delta$ lacX74 mcrA $\Delta$ (mrr-hsdRMS-mcrBC) recA1 endA1 araD139 $\Delta$ (ara, leu)7697 galU galK $\lambda^-$ rpsL nupG		Epicentre
<b>JM109</b>	e14 <sup>-</sup> (mcrA <sup>-</sup> ) recA1 endA1 gyrA96 thi-1 hsdR17(rk <sup>-</sup> mk <sup>+</sup> ) supE44 $\Delta$ (lac-proAB) [F' traD36 proAB lacI <sup>q</sup> Z $\Delta$ M15]	A. nal	Appendice
<b>S17.1(pir)</b>	thi pro hsdR <sup>-</sup> hsdM <sup>+</sup> recA RP4[trp sm] 2-Tc::Mu-kn::Tn7 res <sup>-</sup> mod <sup>+</sup> $\lambda$ pir	Sm, Sp	Simon, 1983
<b>CC118 <i>immE3</i></b>	araD139 $\Delta$ (ara, leu)7697 $\Delta$ lacX74 phoAD20 galE galK thi rpsE rpoB argE(am) recA1 immE3	Kn	Diaz <i>et al.</i> , 1994
<b>TE2680</b>	F <sup>-</sup> $\lambda^-$ IN(rrnD-rrnE)1 $\Delta$ lacX74 rpsL galK2 recD1903::Tn10d-Tet trpDC700::putA1303::[cm <sup>r</sup> -lac]	Cm, Tc	Elliott, 1992
<b>TELI</b>	F <sup>-</sup> $\lambda^-$ IN(rrnD-rrnE)1 $\Delta$ lacX74 rpsL galK2 recD1903::Tn10d-Tet trpDC700::putA1303::[kn <sup>r</sup> -lacI <sup>q</sup> -immE3-lac]	Kn, Tc	cette étude

A.nal : acide nalidixique, Amp : ampicilline, Cm : chloramphénicol, Kn : kanamycine, Sm: streptomycine, Sp: spectinomycine, Tc : tétracycline,

2004). Le promoteur  $P_{bad}$  régulé par AraC a également été testé, ainsi que le promoteur hybride  $P_{lac/araI}$  régulé à la fois par LacI et AraC. Ce dernier n'autoriserait selon les auteurs qu'une très faible expression basale (Lutz & Bujard, 1997).

Ces tests ont été réalisés avec les gènes *relF* et *colE3*, deux gènes toxiques entraînant la mort de la cellule hôte selon des modes d'action différents. Le gène *relF* code une porine s'ancrant dans la membrane, entraînant l'abolition du potentiel membranaire, ce qui a pour conséquence la mort de la cellule hôte (Molin *et al.*, 1993). Le gène *colE3* code une colicine à activité ribonucléasique. La protéine ColE3 clive la terminaison 3' de l'ARNr16S. Le site de liaison à l'ARNm est alors éliminé, ce qui entraîne l'inhibition de la synthèse protéique et aboutit à la mort de la cellule hôte (Senior & Holland, 1971).

Après comparaison de l'efficacité des différents systèmes construits, un système d'expression inductible efficace a été choisi pour chacun des gènes toxiques. Le gène *colE3* s'étant révélé extrêmement toxique pour la cellule hôte, un nouveau système de contrôle de son expression a été développé.

## **2 MATERIELS ET METHODES**

### **2.1 Bactéries et plasmides**

Les souches de *E. coli* et les plasmides construits ou utilisés sont répertoriés dans le **Tableau 1** et le **Tableau 2**. Les souches ont été conservées dans du glycérol (Euromedex, Souffelweyersheim, France) 25 % en concentration finale et stockées à -80°C. Les étalements et cultures liquides ont été effectués dans le milieu nutritif Luria Bertani modifié (LBm) (**Annexe 1**) contenant ou non de l'agar à 10 g.L<sup>-1</sup> et en présence de la pression de sélection adéquate. Les antibiotiques ont été ajoutés dans le milieu pour une concentration finale suivante : ampicilline 100 µg.mL<sup>-1</sup> (Euromedex), tétracycline 20 µg.mL<sup>-1</sup> (Sigma-Aldrich, Saint-Quentin Fallavier, France), chloramphénicol 25 µg.mL<sup>-1</sup> (Euromedex), streptomycine 50 µg.mL<sup>-1</sup> (Sigma-Aldrich), kanamycine 50 µg.mL<sup>-1</sup> (Euromedex). L'ampicilline a été dans certains cas été remplacé par de la carbénicilline (100 µg.mL<sup>-1</sup>) (Sigma-Aldrich). Ce dernier, plus résistant à l'hydrolyse par la β-lactamase, diminue l'apparition de colonies satellites. Du D(+) glucose (Carlo Erba Réactifs, Val de Reuil, France) à 2 g.L<sup>-1</sup>, du L(+) arabinose (Carl Roth, Lauterbourg, France) à 10 g.L<sup>-1</sup> et de l'IPTG (Euromedex) à 500 µg.mL<sup>-1</sup> ont été ajoutés dans le milieu quand précisé. Toutes les solutions dissoutes dans l'eau ont été stérilisées par passage sur filtre 0,2 µm (Millipore, Saint Quentin en Yvelines, France).



**Tableau 2 : Liste et caractéristiques des plasmides utilisés et construits**

Plasmides	Génotype	Origine réplication	Utilisation	Référence
pMMB190	<i>amp<sup>r</sup>, lacI<sup>q</sup>, P<sub>taclacUV5</sub>::lacZα</i>	incQ	Vecteur expression	Morales <i>et al.</i> , 1991
pMR32	<i>amp<sup>r</sup>, lacI<sup>q</sup>, P<sub>taclacUV5</sub>::relF</i>	incQ	Construction toxique	cette étude
pMR32Tc	<i>amp<sup>r</sup>, tc<sup>r</sup>, lacI<sup>q</sup>, P<sub>taclacUV5</sub>::relF</i>	incQ	Construction toxique	cette étude
pMRTc20	<i>amp<sup>r</sup>, tc<sup>r</sup>, lacI<sup>q</sup>, P<sub>taclacUV5</sub>::lacZα, P<sub>A10304</sub>::relF</i>	incQ	Construction toxique	cette étude
pFEB98	<i>amp<sup>r</sup>, lacI<sup>q</sup>, 2 x (P<sub>A1-O3/O4</sub>::relF)</i>	incQ	Construction toxique	cette étude
pMC	<i>amp<sup>r</sup>, cm<sup>r</sup>, lacI<sup>q</sup>, P<sub>A1-O3/O4</sub>::colE3</i>	incQ	Construction toxique	cette étude
pUCTc	<i>amp<sup>r</sup>, tc<sup>r</sup>, lacI<sup>q</sup>, P<sub>lac</sub>::lacZα</i>	pMB1 modifié	Source du gène de résistance à la tétracycline	Mavingui (non publié)
pJBA25	<i>amp<sup>r</sup>, P<sub>lac</sub>::gfp</i>	colE1	Source du gène <i>gfp</i>	Moller <i>et al.</i> , 1998
pSK360	<i>amp<sup>r</sup>, tc<sup>r</sup>, 2 x (P<sub>A1-O3/O4</sub>::relF)</i>	pMB1	Source gène suicide <i>relF</i>	Knudsen <i>et al.</i> , 1995
pColCA38	<i>colE3, immE3, immE8</i>	?	Source gène toxique <i>colE3</i> , gène antitoxique <i>immE3</i>	Masaki & Ohta, 1982
pEDF5	<i>tc<sup>r</sup>, lacI<sup>q</sup>, P<sub>lac</sub>::colE3</i>	incQ	Source gène suicide <i>colE3</i>	Diaz <i>et al.</i> , 1994
pMMS3	<i>amp<sup>r</sup>, sm<sup>r</sup>/sp<sup>r</sup>, cm<sup>r</sup>, P<sub>A1-O3/O4</sub>::colE3</i>	R6K	Source gène suicide <i>colE3</i>	Munthali <i>et al.</i> , 1996
pBAD24	<i>amp<sup>r</sup>, araC, P<sub>bad</sub></i>	pMB1	Vecteur expression	Guzman <i>et al.</i> , 1995
pB24G	<i>amp<sup>r</sup>, araC, P<sub>bad</sub>::gfp</i>	pMB1	Témoin expression	cette étude
pBAD33	<i>cm<sup>r</sup>, araC, P<sub>bad</sub></i>	p15A	Vecteur expression	Guzman <i>et al.</i> , 1995
pB33G	<i>cm<sup>r</sup>, araC, P<sub>bad</sub>::gfp</i>	p15A	Témoin expression	cette étude
pB33R	<i>cm<sup>r</sup>, araC, P<sub>bad</sub>::relF</i>	p15A	Construction toxique	cette étude
pB33C3	<i>cm<sup>r</sup>, araC, P<sub>bad</sub>::colE3</i>	p15A	Construction toxique	cette étude
pB33C3R	<i>cm<sup>r</sup>, tc<sup>r</sup>, araC, P<sub>bad</sub>::colE3, P<sub>A1-O3/O4</sub>::relF</i>	p15A	Construction toxique	cette étude
pZS24MCS1	<i>kn<sup>r</sup>, P<sub>lac/ara1</sub></i>	pSC101 modifiée	Vecteur expression	Lutz <i>et al.</i> , 1997
pZSCm	<i>cm<sup>r</sup>, P<sub>lac/ara1</sub></i>	pSC101 modifiée	Vecteur expression	cette étude
pZSCmC3	<i>cm<sup>r</sup>, P<sub>lac/ara1</sub>::colE3</i>	pSC101 modifiée	Construction toxique	cette étude
pRS551	<i>amp<sup>r</sup>, kn<sup>r</sup>, lacZ, lacY, lacA</i>	pMB1	Vecteur intégration chromosomique	Simons <i>et al.</i> , 1987
p2753	<i>amp<sup>r</sup>, kn<sup>r</sup>, lacI<sup>q</sup>, ΔlacY, lacA</i>	pMB1	Vecteur intégration chromosomique	cette étude
pRSLI1	<i>amp<sup>r</sup>, kn<sup>r</sup>, P<sub>bla</sub>::immE3, lacI<sup>q</sup>, ΔlacY, lacA</i>	pMB1	Vecteur intégration chromosomique	cette étude
P2774	<i>amp<sup>r</sup>, kn<sup>r</sup>, lacI<sup>q</sup>, ΔlacZ, lacY, lacA</i>	pMB1	Vecteur intégration chromosomique	cette étude
pRSLI2	<i>amp<sup>r</sup>, kn<sup>r</sup>, P<sub>bla</sub>::immE3, lacI<sup>q</sup>, ΔlacZ, lacY, lacA</i>	pMB1	Vecteur intégration chromosomique	cette étude
pBR328	<i>amp<sup>r</sup>, cm<sup>r</sup>, tc<sup>r</sup></i>	pMB1	Source <i>P<sub>bla</sub></i>	Gilbert, 1991(non publié)

pMMB190 et dérivés, pBAD 24 et dérivés, pBAD33 et dérivés, pZS24MCS1 et dérivés, pRS551 et dérivés, autres plasmides

Les plasmides ont été extraits avec les kits commerciaux Qiagen Plasmid Mini, Midi et Maxi (Courtaboeuf, France), Macherey-Nagel NucleoBond PC100 (Hoerdts, France), et Promega Wizard plus Miniprep (Charbonnières, France) selon les recommandations du fournisseur. L'extraction et purification de plasmide sur gradient de chlorure de césium ont également été testées (**Annexe 2**). La qualité et la concentration de l'extrait plasmidique ont été vérifiées par dosage au spectrophotomètre ND-1000 (NanoDrop Technologies, Wilmington, United States of America). Le plasmide extrait a également été contrôlé par digestion enzymatique.

## 2.2 Digestions enzymatiques

Des mélanges réactionnels de 25  $\mu\text{L}$  ont été réalisés avec par réaction : 100 à 500 ng d'ADN suivant les cas, le tampon de digestion adapté (Fermentas, Mundolsheim, France), (les) l'enzyme(s) de restriction (Fermentas), le tout complété avec de l'eau stérile (Aguettant, Lyon, France). Les mélanges ont été incubés à 37°C entre 1 h et 2 h ou plus suivant les enzymes utilisées avant inactivation de l'enzyme selon les recommandations du fabricant. Les produits de digestion ont été vérifiés par migration sur gel d'agarose (InvitroGen, Cergy Pontoise, France) de 1 % (10  $\text{g.L}^{-1}$ ), à 2 % (20  $\text{g.L}^{-1}$ ) d'agarose suivant la taille des fragments à visualiser, contenant du Bromure d'Ethyidium (BET) (Carlo Erba Reactifs, Val de Reuil, France) à 0,5  $\mu\text{g.mL}^{-1}$  en concentration finale. Du bleu de bromophénol a été ajouté à l'ADN lors du dépôt dans les puits du gel selon un ratio 1 :5. La taille des fragments d'ADN a été contrôlée par la migration conjointe du marqueur de taille 1 Kb plus (1Kb+) (InvitroGen, Cergy Pontoise). La migration a été effectuée dans les cuves RunOne (EmbiTec, La Jolla, United States of America) sous une tension de 50 à 100 V dans du tampon TBE (Euromedex, Souffelweyersheim, France) 0,5 X. Les gels ont été visualisés sous Ultra Violets (UVs) avec le système GelDoc-it (UVP, Cambridge, United Kingdom).

## 2.3 Amplification de l'ADN par PCR

Les réactions enzymatiques PCR ont été réalisées avec la Taq polymérase Invitrogen (Cergy Pontoise, France). Les mélanges réactionnels ont été préparés sous hotte irradiée 10 min aux rayons UVs. Les composants du mélange réactionnel ont été ajoutés à la concentration finale suivante : tampon réactionnel à 1 X, solution de chlorure de magnésium ( $\text{MgCl}_2$ ) à 1,5 mM, mélange de dNTPs (fournis avec l'enzyme Taq) à 200  $\mu\text{M}$ . Les amorces sens (Fw) et antisens (Rv), définies sous oligo 6.1 et commandées chez Invitrogen, ont été

**Tableau 3 : Liste des amorces, de leurs caractéristiques et des conditions d'amplification utilisées**

Amorce	Séquence en 5'-	ADN cible	Gène amplifié	Site restriction	Mésappariements <sup>b</sup>	Tm <sup>a</sup> (°C)
<b>immE3pvuIFw4</b>	TATCCGATCGAGGAAGTTATGGGACTTAAATTG	pColCA38	<i>immE3</i>	PvuI	5	66
<b>immE3pstIRv5</b>	AGCCCTGCAGTCAACTCATCCCTGATAATATTT	pColCA38	<i>immE3</i>	PstI	5	66
<b>p3imme3bamHIFw3</b>	TGAGATTATCAAAAAGGATCCTCAC	pBR328-immE3	<i>P3 ::immE3</i>	BamHI	1	55
<b>p3imme3ecoRIRv1</b>	TTCCGTCACAGGTATGAATTCGAAG	pBR328-immE3	<i>P3 ::immE3</i>	EcoRI	2	55
<b>ZS24Aat2Fw2</b>	CCGACGCTCTGTGTGGAATTGTGAG	pZS24MCS1	<i>P<sub>lac/ara1</sub>, ori pSC101</i>	AatII	0	62
<b>ZS24XbaIRv2</b>	CTGGGGTTCTAGAGCTCGCTTGGACT	pZS24MCS1	<i>P<sub>lac/ara1</sub>, ori pSC101</i>	XbaI	1	62
<b>CmaatIIFw5</b>	GGCAGCATCACCCGACGTCCTTTG	pB33C3	<i>Cm<sup>R</sup></i>	AatII	2	62
<b>CmxbaIRv2</b>	CCTCAGGCATCTCTAGAAGCACACGGTCA	pB33C3	<i>Cm<sup>R</sup></i>	XbaI	2	62
<b>colE3KpnIFw5</b>	ACAGGTACCATCGTGTAATCAATTGTTTAAAC	pColCA38	<i>colE3</i>	KpnI	4	60
<b>colE3MluIRv5</b>	TAAACGCGTCTCAAAGATATTCTTGATATTTC	pColCA38	<i>colE3</i>	MluI	4	60

Les sites de restriction sont indiqués sur la séquence par un soulignement

<sup>a</sup>Tm : température d'hybridation a été définie expérimentalement

<sup>b</sup> Ajout de mésappariements par rapport à la séquence nucléotidique de l'ADN cible

ajoutées à une concentration finale de 0,5  $\mu\text{M}$ . L'enzyme Taq polymérase a été ajoutée à raison de 0,025  $\text{U} \cdot \mu\text{L}^{-1}$  et l'échantillon d'ADN à amplifier (plasmide) à 0,004  $\text{ng} \cdot \mu\text{L}^{-1}$ .

Les conditions d'amplification ont été définies avec l'appareil thermocycleur Biometra T-gradient (Labgene Scientific Instruments, Archamps, France) selon un gradient de température. Les amplifications routinières ont été effectuées dans le thermocycleur Biometra T-personal. Les amorces définies ainsi que la température d'hybridation  $T_m$  utilisée sont répertoriées dans le **Tableau 3**. Les amplifications ont été effectuées par une dénaturation initiale à 95°C 5 min, suivi de 30 à 35 cycles constitués chacun des étapes de dénaturation à 95°C 1 min, d'hybridation à  $T_m$  1 min, puis d'élongation 72°C 1 min.kb<sup>-1</sup>. Une élongation finale de 72°C a été effectuée pendant 10 min. Les produits PCR ont été contrôlés par migration sur gel d'agarose selon les conditions indiquées ci-dessus.

## 2.4 Clonages

Les fragments d'ADN provenant de la restriction de plasmides ou de produits d'amplification digérés ou non ont été utilisés pour les réactions de ligation. Ces derniers ont été purifiés à l'aide des kits commerciaux Qiaquick PCR purification and gel extraction (Qiagen, Courtaboeuf, France), GFX PCR DNA and gel band purification (Amersham, Buckinghamshire, United Kingdom) ou NucleoSpin ExtractII (Macherey-Nagel, Hoerd, France) soit directement sur colonne, soit à partir d'un gel (**Annexe 3**). Lorsque précisé, des bouts francs ont été générés sur les extrémités cohésives des fragments d'ADN par l'utilisation de l'enzyme Klenow Fragments (Roche Diagnostics, Meylan, France). Le mélange réactionnel et les conditions d'incubation utilisées ont été celles recommandées par le fabricant, puis l'ADN a été purifié sur colonne commerciale. La déphosphorylation des vecteurs a été effectuée par traitement avec l'enzyme Phosphatase Alcaline Shrimp (SAP) (Roche Diagnostics) selon le protocole répertorié en **Annexe 4**. L'ADN a été ensuite purifié sur colonne commerciale. Les fragments d'ADN utilisés pour le clonage ont été quantifiés sur gel, en présence des marqueurs de masse MassRuler High Range et / ou MassRuler Low Range (Fermentas, Mundolsheim, France) suivant la taille des fragments attendue. Les fragments d'ADN, insert(s) et vecteur ayant leurs deux extrémités franches ont été mélangés selon un ratio molaire insert : vecteur 5 :1. Dans les autres cas inserts et vecteurs ont été mélangés selon un ratio molaire 1 :1 s'ils avaient une taille identique, 3 :1 si l'insert était inférieur en taille au vecteur. Un mélange réactionnel de ligation de 20  $\mu\text{L}$  a été composé du mélange insert : vecteur, de 2  $\mu\text{L}$  de tampon adapté et de minimum 1  $\mu\text{L}$  de l'enzyme T4 ligase à 1  $\text{U} \cdot \mu\text{L}^{-1}$  (Roche, Meylan), le tout complété avec de l'eau stérile (Aguettant, Lyon,

France). Le mélange a été incubé 1 nuit à 16°C, 10 min à 70°C, puis conservé sur glace avant transformation.

## 2.5 Préparation des cellules pour l'électroporation

Les souches *E. coli* ont été rendues électrocompétentes suivant le mode opératoire suivant. Une colonie isolée fraîche a été inoculée en milieu liquide en présence de(s) pression(s) de sélection adéquate(s). La préculture a été incubée 1 nuit à 37°C sous agitation orbitale de 150 rpm, puis a été diluée au 1/100<sup>ème</sup> dans du LBm contenant le(s) antibiotique(s) adéquat(s). La culture a été incubée à 37°C sous agitation de 150 rpm jusqu'à une densité optique (DO) mesurée à 600 nm au spectrophotomètre Biophotometer (Eppendorf, Le Pecq, France) de 0,6 maximum. A partir de cette étape, les cellules bactériennes ont été maintenues à une température maximale de 4°C sur glace ou dans des appareils réfrigérants. La culture bactérienne a été centrifugée à 4 000 g 10 min à 4°C, le surnageant a été éliminé, et les cellules sédimentées ont été lavées dans du glycérol 10 % pré refroidi à 4°C. Les étapes de centrifugation / lavage ont été effectuées 4 fois, avec un volume de glycérol 10 % froid de 1, 1/2, 1/5<sup>ème</sup> puis 1/50<sup>ème</sup> du volume initial de la culture bactérienne. Les cellules ont été reprises dans un volume final de glycérol froid 10 % correspondant à 1/100<sup>ème</sup> du volume initial, puis utilisées extemporanément ou aliquotées et stockées à -80°C.

## 2.6 Transformation

L'ADN a été transformé dans la souche *E. coli* DH5 $\alpha$  chimiocompétente (InvitroGen, Cergy Pontoise, France) par choc thermique. Pour ce faire, 5  $\mu$ L de la solution d'ADN (5 à 200 ng ADN) ont été mis en présence de 50  $\mu$ L de cellules. Le choc a été effectué selon les recommandations du fabricant. L'ADN a été transformé dans les autres souches par électroporation. Les souches ont été préparées selon le protocole indiqué ci-dessus, et utilisées soit extemporanément, soit à partir des stocks à -80°C après décongélation des cellules 30 min environ sur glace. Dans une cuve d'électroporation de 2 mm (BTX, MP Biomedicals, Illkirch, France) pré-refroidie sur glace, 2  $\mu$ L de solution ADN ont été mis en présence de 50  $\mu$ L de cellules et le volume a été ajusté avec 3  $\mu$ L d'eau stérile (Aguettant, Lyon, France). Un choc électrique de 2,5 kV.cm<sup>-1</sup>, 200  $\Omega$  et 25  $\mu$ F a été appliqué avec GenepulserII (Bio-Rad, Marne La Coquette, France). Après choc (électrique ou chimique), du milieu LBm a été ajouté en quantité suffisante pour (QSP) 1ml, puis les cellules ont été incubées de 1 h à 2 h à 37°C sous agitation douce (<150 rpm). La suspension bactérienne a été diluée en série au 1/10<sup>ème</sup> dans

une solution de chlorure de sodium (NaCl) à 8 g.L<sup>-1</sup>. Les dilutions 10<sup>0</sup>, 10<sup>-1</sup>, 10<sup>-2</sup> ont été étalées sur milieu gélosé sélectif adéquat pour dénombrer les transformants, les dilutions 10<sup>-4</sup>, 10<sup>-5</sup> et 10<sup>-6</sup> ont été étalées sur milieu gélosé non sélectif pour dénombrer les cellules réceptrices. Après incubation 1 nuit à 37°C, les colonies ont été dénombrées.

Une fréquence de transformation a été estimée en divisant le nombre d'Unités Formant Colonies par mL (UFC.mL<sup>-1</sup>) de transformants sur le nombre d'UFC.mL<sup>-1</sup> de réceptrices. Une fréquence de transformation par µg d'ADN et par mol d'ADN a également été calculée. La préparation de stocks de cellules électrocompétentes, bien que standardisée, donne lieu à une transformabilité des cellules variable d'un stock à l'autre. Une fréquence de transformation témoin a donc été estimée par la transformation d'un plasmide témoin pour chaque nouveau stock de cellules préparées.

## 2.7 Modification chromosomique

La cassette [*lacI<sup>q</sup>*-*kn'*-P3::*immE3*] a été insérée dans la souche *E. coli* TE2680 par clonage de la cassette dans le plasmide pRS551, puis linéarisation et transformation du plasmide obtenu (pRSLI2) dans la souche TE2680 selon le protocole donné par les auteurs (Simons *et al.*, 1987).

## 2.8 Criblage phénotypique

### Par transformation :

La fonctionnalité des gènes toxiques a été vérifiée par transformation des plasmides portant ces gènes dans une souche ne portant ni gène d'immunité, ni gènes répresseurs *lacI<sup>q</sup>* ou *araC* suivant les promoteurs contrôlant leur expression. La fréquence de transformation est calculée et comparée à la fréquence de transformation d'un plasmide témoin. La fonctionnalité du gène *lacI<sup>q</sup>* intégré au niveau chromosomique a également été vérifiée par transformation d'un plasmide contenant un gène toxique (*relF*) sous contrôle d'un promoteur dérivé de P<sub>lac</sub> (P<sub>A1-O3O4</sub>). La fréquence de transformation de ce plasmide a été calculée et comparée à la fréquence de transformation d'un plasmide témoin. Un contrôle ultérieur a été effectué par induction de l'expression du gène toxique du transformant obtenu en milieu liquide en présence d'IPTG.

Par induction en milieu liquide :

L'insertion du gène toxique et sa fonctionnalité ont été vérifiées en milieu liquide par criblage des transformants après induction de l'expression. Les transformants ont été inoculés en microplaque 96 puits cellstar (Greiner Bio-one, Courtaboeuf, France) contenant 100 µL de LBm et les antibiotiques (AB) adéquats. Les microplaques ont été incubées 4 h à 37°C avec agitation puis les clones ont été transférés individuellement par réplique dans des microplaques contenant (a) 100 µL LBm liquide + AB (b) 100 µL LBm + AB + inducteur (s) de l'expression toxique. De l'IPTG a été utilisé pour induire l'expression des gènes toxiques sous contrôle du système régulateur dérivé de l'opéron lactose. Du L(+) arabinose a été utilisé pour induire l'expression des gènes toxiques sous contrôle du système régulateur dérivé de l'opéron arabinose. Les plasmides des transformants présentant le phénotype recherché ont été extraits à l'aide de kits commerciaux et leur construction génétique contrôlée par restriction enzymatique.

Par induction en milieu solide :

La fonctionnalité des constructions toxiques construites a été contrôlée suivant les cas par striation, repiquage ou étalement des transformants sur milieu solide sélectif (a) sans inducteur ou (b) avec inducteur selon des concentrations identiques à celles utilisées en milieu liquide et incubation 1 nuit à 37°C. La fonctionnalité des gènes toxiques a été vérifiée par l'absence de développement de la souche testée sur milieu avec inducteur par rapport au milieu sans inducteur. Pour la vérification de l'expression du gène marqueur *lacZ*, du X-gal (5-bromo-4-chloro-3-indolyl-β-D-Galactopyranoside) (Euromedex, Souffelweyersheim, France) a été ajouté dans les milieux avec et sans inducteur. Le niveau d'expression a été observé visuellement par l'intensité de la coloration bleue des colonies. Pour le gène *gfp*, le niveau d'expression a été observé sous microscope AxioSkop Zeiss (Oberkochen, Allemagne) équipé d'un filtre set 10 (filtre GFP) (excitation : bande passante (BP) 450/490nm, émission : BP 510nm) par l'intensité de la coloration verte des colonies. Dans le cas du gène *immE3*, son insertion et sa fonctionnalité ont été vérifiées par repiquage des cellules à tester sur milieu contenant ou non de la colicine. Le protocole d'extraction de la colicine est indiqué en **Annexe 5**. Le volume d'extrait de colicine à utiliser par volume de milieu a été déterminé au préalable sur des souches *E. coli* témoins. La fonctionnalité du gène *immE3* a été vérifiée par le développement de la souche en présence et absence de colicine, par rapport au témoin négatif (souche avant insertion chromosomique qui ne se développe pas uniquement sur milieu contenant de la colicine).

## 2.9 Test d'efficacité

Une colonie fraîche a été inoculée en milieu liquide contenant les antibiotiques adéquats. Après 12 h d'incubation à 37°C sous agitation de 150 rpm, la culture bactérienne a été diluée en série au 1/10<sup>ème</sup> dans une solution de NaCl à 8 g.L<sup>-1</sup> jusqu'à la dilution 10<sup>-6</sup>. Cent µL de chacune des dilutions ont été étalés sur boîte de Petri contenant du milieu gélosé sélectif avec ou sans inducteur (IPTG à 500 µg.mL<sup>-1</sup> à 1 mg.mL<sup>-1</sup> ou arabinose à 10 g.L<sup>-1</sup>). Les boîtes de Petri ont été incubées 1 nuit à 37°C, les colonies apparues (UFC) ont été dénombrées. Un taux d'échappement a été calculé en divisant le nombre d'UFC.mL<sup>-1</sup> se développant sur milieu avec inducteur par celui des UFC.mL<sup>-1</sup> se développant sur milieu sans inducteur. Pour augmenter le nombre d'UFC.mL<sup>-1</sup> soumis au test d'efficacité, la culture bactérienne a été concentrée par centrifugation, élimination du surnageant et remise en suspension dans un volume au moins 5 fois plus faible.

## 2.10 Cinétiques de croissance

Une colonie fraîche a été inoculée en milieu liquide en présence des antibiotiques adéquats. Après 12 h d'incubation à 37°C sous agitation de 150 rpm, la DO a été mesurée à 600 nm. La DO<sub>600nm</sub> des suspensions bactériennes a été ajustée à 0,3 avec du milieu LBm. Cent µL de cette suspension ont été inoculés dans 100 mL de LBm contenant la pression de sélection adéquate. La DO<sub>600nm</sub> a été mesurée toutes les heures pendant 8 h. Des échantillons de 1 mL prélevés toutes les 2 h ont été utilisés pour effectuer un dénombrement des cellules bactériennes vivantes et cultivables.

# 3 RESULTATS ET DISCUSSION

## 3.1 Contrôles et optimisations avant clonage de gènes toxiques

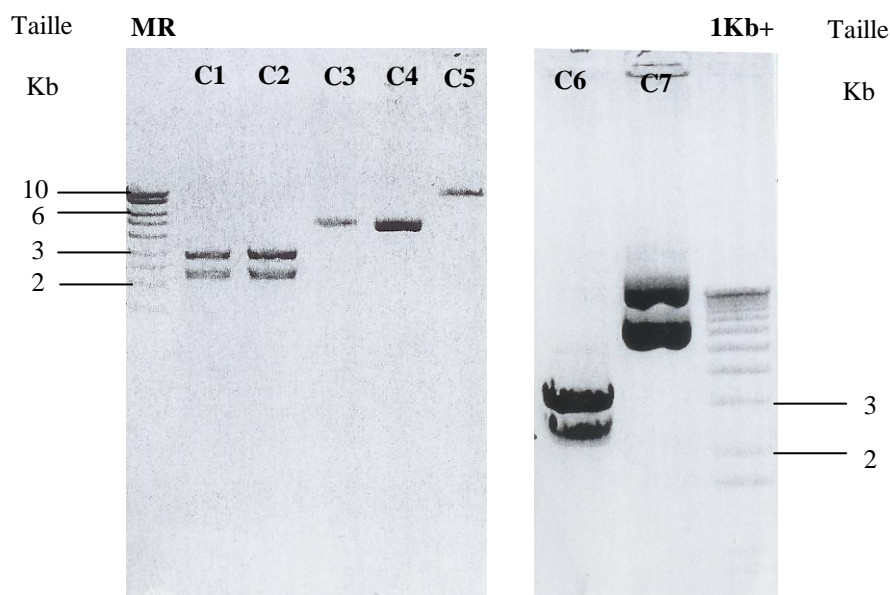
### 3.1.1 Transformabilité des cellules électrocompétentes :

La transformabilité des cellules électrocompétentes a été vérifiée par la transformation d'un plasmide témoin. Une plus haute fréquence de transformation de plasmide pour des cellules préparées extemporanément a été constatée par rapport à celle obtenue pour les mêmes cellules stockées à -80°C (**Tableau 4**). Les fréquences d'obtention des transformants comportant les constructions toxiques attendues étant parfois très faibles (voir ci-après), les transformations avec des produits de ligation contenant des gènes toxiques ont été systématiquement réalisées avec des cellules électrocompétentes préparées extemporanément.



**Tableau 4 : Fréquence de transformation de plasmides dans la souche *E. coli* JM109 en fonction des conditions d'utilisation des souches préparées électrocompétentes**

Souche <i>E.coli</i>	plasmide	T	T. $\mu\text{g}^{-1}$ ADN	T.mol <sup>-1</sup> ADN	Condition d'utilisation
JM109	<b>pBAD24</b>	1,60E-03	1,07E-02	4,84E-05	extemporanément
JM109	<b>pBAD24</b>	1,77E-04	1,18E-03	5,35E-06	après stockage -80°C
JM109	<b>pMMB190</b>	1,37E-04	6,86E-04	6,22E-06	extemporanément
JM109	<b>pMMB190</b>	2,39E-05	1,20E-04	1,08E-06	après stockage -80°C



**Figure 4 : Comparaison des profils de digestion entre les plasmides pMR32, pUCTc, et les plasmides extraits de transformants.**

Les plasmides extraits des clones *E. coli* JM109 dans lesquels ont été introduits les produits de ligation de pMR32 et de l'insert contenant le gène *tc<sup>r</sup>* (provenant de pUCTc) ont été digérés par l'enzyme de restriction HindIII, et comparé au profil de digestion de pUCTc et pMR32 digérés par la même enzyme. **C1**: « pMRTc » clone1 digéré par HindIII, **C2** : « pMRTc » clone 2 digéré par HindIII, **C3** : « pMRTc » clone1, **C4** : « pMRTc » clone 2, **C5** : pMR32 digéré par HindIII, **C6** : pUCTc digéré par HindIII, **C7** : pUCTc. **MR** (MassRuler high range) et **1Kb+** les marqueurs de taille.

Pour chaque transformation réalisée, un plasmide témoin est utilisé pour évaluer la transformabilité de chaque stock de cellules électrocompétentes. Les fréquences de transformation T obtenues pour ces plasmides témoins nous ont servi de référence interne pour évaluer les fréquences de transformation obtenues pour chaque produit de ligation. Une fréquence de transformation normalisée TN est alors calculée :

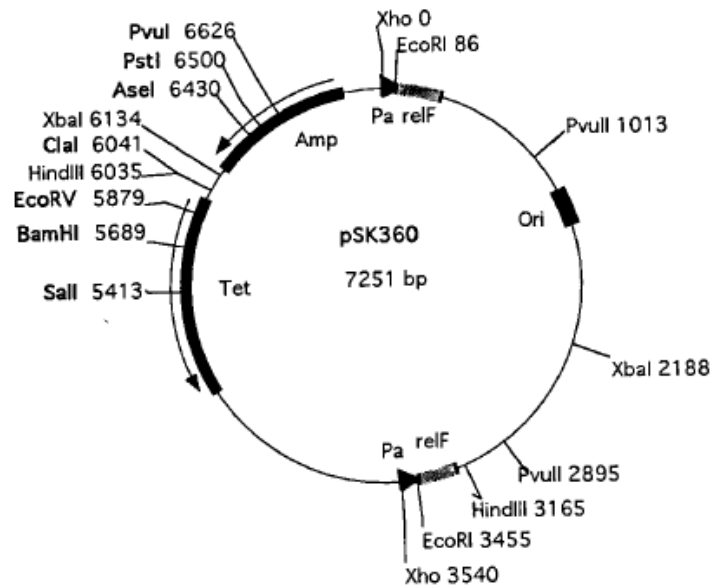
$$TN = T(\text{produit de ligation}) / T(\text{plasmide témoin})$$

Les plasmides témoins ont été choisis en fonction des vecteurs utilisés pour réaliser le clonage. Les plasmides pMMB190 et pBAD33 ont ainsi été utilisés pour évaluer la fréquence de transformation des constructions réalisées dans ces deux vecteurs. Ce choix permet (1) de vérifier que le vecteur utilisé est capable de se répliquer de manière autonome dans la souche hôte utilisée, (2) de comparer la transformabilité normalisée TN de tous les essais de construction dérivés d'un même vecteur.

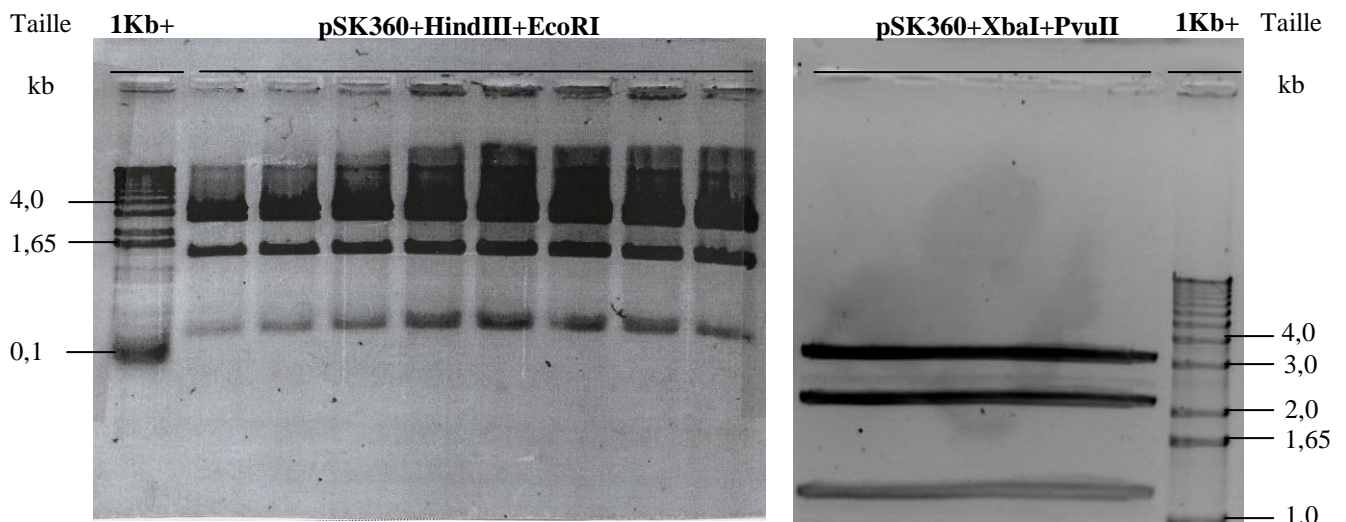
### 3.1.2 Optimisations des techniques de clonage :

Les premiers essais de construction moléculaire incluant des gènes dont les fonctions sont toxiques pour la cellule se sont soldés par des échecs : soit aucun clone n'était détecté, soit les colonies obtenues sur milieu sélectif possédaient un plasmide différent de celui attendu. Ainsi pour la construction du plasmide pMR32Tc, nous avons digéré le plasmide pUCTc par l'enzyme HindIII, puis purifié sur gel le fragment d'ADN correspondant au fragment contenant le gène de résistance à la tétracycline *tc<sup>r</sup>*. Le vecteur pMR32 a été digéré par la même enzyme, puis déphosphorylé pour éviter sa recircularisation et ainsi favoriser sa ligation avec l'insert. Après transformation du produit de ligation dans une souche *E. coli* et sélection sur tétracycline, les plasmides extraits des transformants obtenus ont montré un profil de restriction similaire à celui de pUCTc (**Figure 4**). Ces résultats peuvent être expliqués par la co-extraction de molécules non digérées du plasmide pUCTc. Leur transformabilité supérieure aux molécules hybrides vecteur : insert assure la production majoritaire de clones dépourvus de la construction attendue.

Ces biais, liés à une digestion imparfaite du plasmide initial, nous ont conduits à la purification des plasmides par ultracentrifugation sur gradient de chlorure de césium pour augmenter la pureté des plasmides extraits. Une amélioration de la digestion des plasmides ainsi purifiés a été constatée (**Figure 5**). Cependant la technique est longue et lourde à appliquer lorsque de nombreux plasmides sont utilisés. Pour éviter l'utilisation de cette méthode de purification, l'emploi d'une seconde enzyme coupant pUCTc en un site différent du gène *tc<sup>r</sup>* permet également de diminuer le nombre de molécules de plasmide non digérées.



(A)



(B)

(C)

**Figure 5 : Influence de la purification des plasmides par ultracentrifugation sur gradient de chlorure de césium sur les réactions de digestion**

(A) carte de restriction de pSK360 (Knusen *et al.*, 1995).

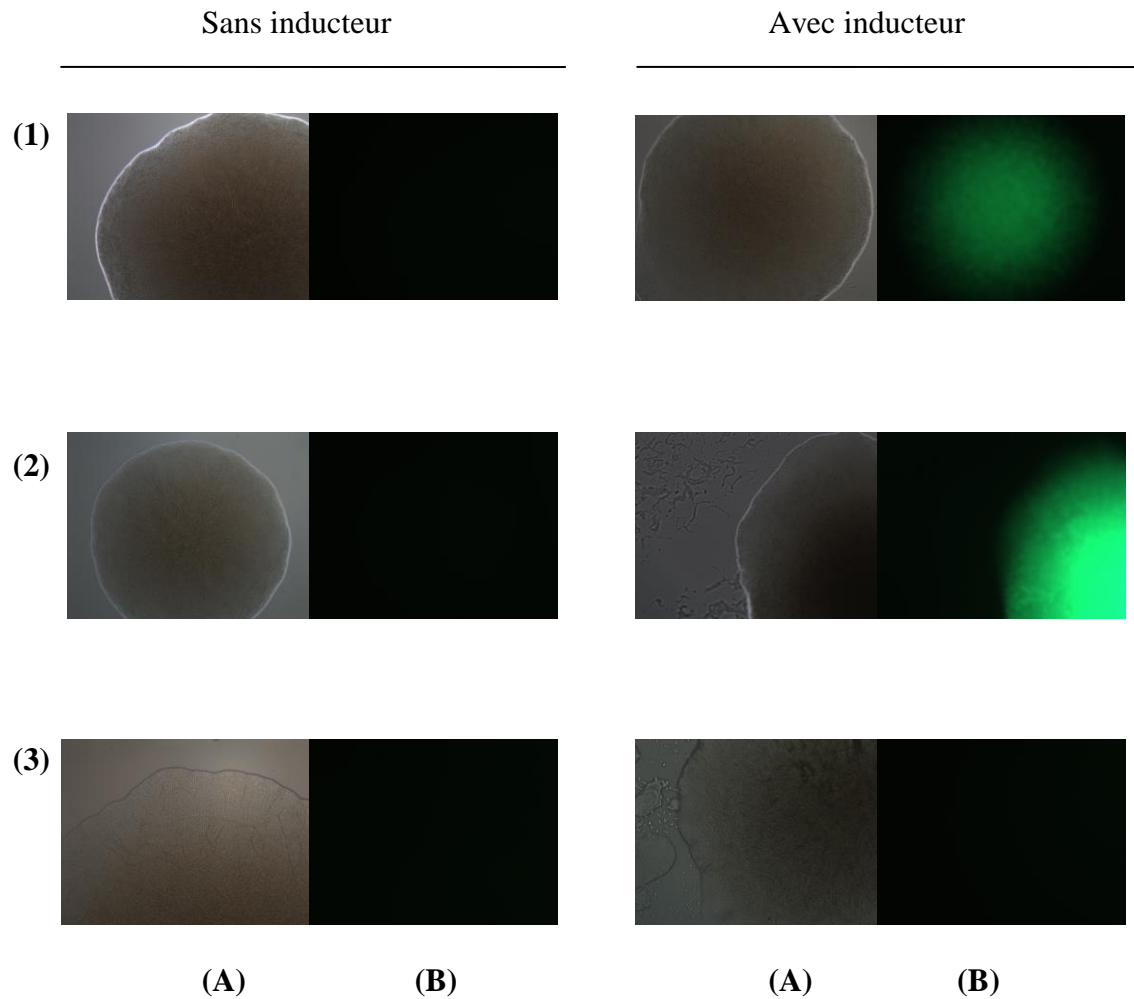
(B) le plasmide pSK360 (3800 ng) extrait par un kit commercial a été digéré par les enzymes de restriction HindIII et EcoRI. (C) le plasmide pSK360 (3600 ng) extrait puis purifié par ultracentrifugation sur gradient de chlorure de césium a été digéré par les enzymes de restriction XbaI et PvuII. La réaction de digestion dans les deux cas a été effectuée dans les mêmes conditions par les mêmes quantités d'enzyme dans un mélange réactionnel final de 200  $\mu$ L. La totalité du mélange réactionnel a été déposé sur gel.

De manière alternative à la digestion, nous avons également expérimenté l'amplification des inserts par PCR. Le plasmide utilisé comme cible a été ajouté dans le mélange réactionnel d'amplification à  $0,004 \text{ ng.}\mu\text{L}^{-1}$  en concentration finale. Cependant les molécules de plasmide matrices ont été re-concentrées lors des étapes de clonage et génèrent un bruit de fond lors du clonage des constructions toxiques.

Lorsque aucune sélection positive n'est apportée par l'insert, la sélection des transformants est effectuée sur le gène de résistance porté par le vecteur. Des transformants ne possédant que le vecteur recircularisé ont cependant été obtenus (faux positifs). Les temps de digestion ont donc été augmentés pour diminuer le nombre de plasmides résiduels non digérés, et deux conditions de déphosphorylation ont été testées pour prévenir la recircularisation des plasmides digérés. Un protocole de déphosphorylation présenté en **Annexe 4** (modifié de celui fourni par les fabricants) a été choisis. Lors de la transformation des produits de ligation vecteur traité : insert, un contrôle de la transformation du produit de ligation du vecteur traité seul a par ailleurs été systématiquement réalisé en parallèle pour évaluer le bruit de fond qu'il constitue. La transformation des produits de ligation de vecteurs seuls digérés par deux enzymes non compatibles et déphosphorylés est généralement inférieure à  $10^{-7}$ , avec une fréquence de transformation normalisée (TN) par mole d'ADN inférieure à  $10^{-2}$ .

### 3.1.3 Vérification d'une expression inductible

Le vecteur pMMB190 à faible nombre de copies par cellule (12 copies chez *E. coli*) (Frey *et al.*, 1992; Morales *et al.*, 1991) porte un promoteur dont l'activité est régulée par la protéine LacI. Celle-ci est produite constitutivement par le gène *lacI<sup>q</sup>* porté par le plasmide. Le gène *lacZ $\alpha$*  présent en aval du promoteur nous a permis de vérifier le contrôle de l'expression du système régulateur porté par ce vecteur dans *E. coli* DH5 $\alpha$ . L' $\alpha$  complémentarité avec le gène *lacZ $\Delta$ M15* présent dans *E. coli* DH5 $\alpha$  permet la production d'une enzyme, la  $\beta$ -galactosidase qui en clivant le X-gal entraîne une coloration bleue visible à l'œil nu. Après transformation de pMMB190 dans *E. coli* DH5 $\alpha$ , les transformants obtenus sont étalés sur milieu contenant du X-gal avec ou sans inducteur. Une coloration bleue des colonies est observée uniquement en présence d'IPTG pour DH5 $\alpha$  (pMMB190), confirmant l'activation de l'expression des gènes sous contrôle du promoteur  $P_{\text{tac} \text{lacUV5}}$  uniquement en présence d'IPTG.



**Figure 6 : Contrôle de l'expression exercé par différentes constructions génétiques**

Les souches *E. coli* JM109 contenant le plasmide (1) pJBA25 (2) pB24G (3) pB33G ont été étalées sur milieu gélosé sélectif contenant ou non l'inducteur de l'expression du gène *gfp*. Les colonies sont observées sous microscope (A) sans filtre (B) avec le filtre GFP. Les photos ont été prises avec l'appareil photo digital AxioCam MRC5.

Un test similaire a été effectué pour des vecteurs portant le promoteur  $P_{bad}$  et le gène *araC*, responsable de la régulation de son expression. Le gène marqueur *gfp* a été cloné en aval du promoteur  $P_{bad}$  dans pBAD24 et pBAD33, vecteurs à respectivement moyen et plus faible nombre de copies par cellule (Guzman *et al.*, 1995). Le contrôle de l'expression des deux plasmides construits, respectivement pB24G et pB33G, a été testé par étalement des souches *E. coli* JM109 (pB24G) et JM109 (pB33G) sur milieu contenant ou non de l'arabinose, inducteur de l'expression. L'expression de la protéine GFP (*green fluorescent protein*) est visualisée sous microscope équipé d'un filtre GFP par l'émission d'une fluorescence ( $\lambda_{\text{émission}} = 510\text{nm}$ ) après excitation de la protéine GFP ( $\lambda_{\text{excitation}} = 488\text{nm}$ ). En absence d'arabinose, aucune cellule verte n'a été observée. En présence d'arabinose, les cellules bactériennes contenant pB24G apparaissent vertes, celles contenant pB33G ne le sont pas (**Figure 6**). Les conditions d'induction utilisées permettent donc une expression forte de la protéine GFP en présence d'arabinose chez pBAD24. Dans le vecteur pBAD33, la quantité de protéines exprimées en présence d'arabinose semble insuffisante pour être détectée avec le matériel de microscopie dont nous disposons. Le bon fonctionnement du système de régulation de l'expression de pBAD33 a néanmoins pu être vérifié ultérieurement lors du clonage du gène *colE3* (voir ci-dessous). La même cassette *gfp* a été clonée dans pBAD24 et pBAD33, la différence majeure entre pBAD33 et pBAD24 tient dans leur réplicon (respectivement p15A et pMB1). La plus faible quantité de protéine GFP exprimée par pB33G pourrait donc être expliquée par son plus faible nombre de copies par cellule. Pour la suite des clonages, pBAD33 a donc été préférentiellement utilisé.

## 3.2 Clonage des gènes toxiques

### 3.2.1 La transformation normalisée, indicateur de la réussite du clonage de gènes toxiques

#### Evaluation de TN pour clonage d'insert non toxique dans les vecteurs

Pour chaque construction génétique réalisée, les conditions de clonage et les fréquences de transformation normalisées sont répertoriées dans le **Tableau 5**. Les fréquences de transformation normalisées peuvent être élevées, certaines pouvant être supérieures à  $10^0$ . Ceci indiquerait non seulement la ligation de toutes les molécules d'inserts dans celles de vecteurs, mais également une meilleure transformabilité du produit de ligation par rapport au vecteur témoin dans certains cas ( $\text{TN.mol}^{-1} > 10^0$ ). Ces données étonnantes peuvent être expliquées par les méthodes de quantification utilisées. Les plasmides sont quantifiés au

**Tableau 5 : Caractéristiques de clonage et fréquences de transformation des constructions génétiques réalisées**

Souche <i>E. coli</i>	Plasmide	Taille du plasmide (pb)	Détails du clonage	T	TN.µg <sup>-1</sup>	TN.mol <sup>-1</sup>	Commentaires
DH5α	<b>pMR32</b>	9 326	<i>relF</i> provenant de pSK360-EcoRI-HinDIII, cloné dans pMMB190-EcoRI-HinDIII	2,22E-03	2,86E+01	2,86E+01	Profil de restriction correct, induction positive
JM109	<b>pMR32Tc</b>	12 306	<i>tc<sup>r</sup></i> provenant pUCTc-HinDIII, cloné dans pMR32-HinDIII-DP	4,90E-07	5,17E-02	5,17E-02	Profil de restriction de pUCTc
JM109	<b>pMR32Tc</b>	12 306	<i>tc<sup>r</sup></i> provenant pUCTc-HinDIII , cloné dans pMR32-HinDIII	1,70E-06	2,76E-02	2,76E-02	Tétracycline résistant, profil de restriction attendu, induction positive
EC100	<b>pMR32RTc</b>	14 400	P <sub>A1-O3/04</sub> :: <i>relF-tc<sup>r</sup></i> provenant pSK360-HinDIII, cloné dans pMR32HinDIII-DP	2,21E-09	4,37E-05	4,37E-05	Tétracycline résistant, induction positive, mais profil de restriction de pSK360
JM109	<b>pMR32RTc</b>	14 400	P <sub>A1-O3/04</sub> :: <i>relF-tc<sup>r</sup></i> provenant pSK360-HinDIII, cloné dans pMR32HinDIII	<b>2,35E-08</b>	<b>2,48E-03</b>	<b>2,48E-03</b>	<b>Inférieur au seuil de détection</b>
DH5αZ1	<b>pMRTC20</b>	12 200	P <sub>A1-O3/04</sub> :: <i>relF-tc<sup>r</sup></i> provenant de pSK360-XbaI-PvuII-K, cloné dans pMMB190-DraI-DP	1,01E-08	9,09E-05	9,09E-05	Profil de restriction correct, induction positive
JM109	<b>pB24G</b>	5400	<i>gfp</i> provenant pJBA25-HinDIII-KpnI, cloné dans pBAD24-HinDIII-KpnI-DP	1,76E-06	9,20E-01	7,80E-01	Pas fluorescence sans arabinose, fluorescence avec arabinose 0,2%
JM109	<b>pB33G</b>	6 200	<i>gfp</i> provenant pJBA25-HinDIII-KpnI, cloné dans pBAD33-HinDIII-KpnI-DP	2,61E-06	1,36E+00	1,36E+00	Pas de fluorescence visualisable avec arabinose 0,2%
JM109	<b>pB33R</b>	5 619	<i>relF-tc<sup>r</sup></i> provenant de pSK360-EcoRI-K-HinDIII, cloné dans pBAD33-SmaI-HinDIII-DP	9,03E-06	4,73E+01	4,73E+01	Profil de restriction correct, induction négative
EC100	<b>pMC</b>	11 100	P <sub>A1-O3/04</sub> :: <i>colE3-cm<sup>r</sup></i> provenant de pMMS3-NotI-K , cloné dans pMMB190-SmaI-DP	4,50E-04	3,56E+01	3,56E+01	Chloramphénicol résistant, profil de restriction correct, induction négative
JM109	<b>pB33C3</b>	7 041	<i>colE3</i> provenant de pEDF5-XbaI-HinDIII, cloné dans pBAD33-XbaI-HinDIII-DP	<b>6,00E-08</b>	<b>5,23E-01</b>	<b>5,23E-01</b>	<b>Inférieur au seuil de détection</b>
CC118immE3	<b>pB33C3</b>	7 041	<i>colE3</i> provenant de pEDF5-XbaI-HinDIII, cloné dans pBAD33-XbaI-HinDIII-DP	8,05E-07	1,98E-01	1,98E-01	Profil de restriction correct, fonctionnalité de <i>colE3</i> vérifiée
DH5αZ1	<b>pZSCm</b>	4301	<i>cm<sup>r</sup></i> amplifié sur pBAD33, digéré par AatII-XbaI, cloné dans pZS24MCS1-AatII-XbaI-DP	4,10E-05	1,25E+00	1,00E+00	Chloramphénicol résistant, profil de restriction correct
DH5αZ1	<b>pZSCmC3</b>	5759	<i>colE3</i> amplifié sur pColCA38, digéré par KpnI-MluI, cloné dans pZSCm-KpnI-MluI-DP	1,22E-04	1,28E-01	1,02E-01	Profil de restriction correct, induction positive

Cases grisées : échecs de construction génétiques, DP : déphosphorylation avec l'enzyme SAP, K : génération de bouts francs avec l'enzyme Klenow Fragments, Le plasmide suivi d'un tiret indique le traitement du plasmide par les enzymes qui suivent, induction positive : mort des cellules en présence d'inducteur, induction négative : pas de mortalité en présence d'inducteur.

spectrophotomètre ND-1000, alors que les molécules d'inserts et de vecteurs sont quantifiées visuellement sur gel d'agarose par analyse comparative avec des marqueurs de masse. La première méthode de quantification donne des valeurs supérieures à celles déduites de la quantification sur gel (données non montrées). En effet, l'absorbance de tous les ADN présents dans la solution, plasmide comme ADN contaminant co-extraît avec les molécules plasmidiques (ADN génomique). La seconde méthode présente l'avantage de quantifier uniquement les fragments d'ADN ciblés. Cette surestimation du plasmide témoin entraîne une sous estimation de la fréquence de transformation de ce plasmide (définie par  $\mu\text{g}$  ou  $\text{mol}$  d'ADN) et donc une surestimation de la fréquence de transformation normalisée du produit de ligation. Les valeurs du tableau sont données à titre indicatif et permettent des comparaisons relatives, les mêmes méthodes de quantification ayant en effet toujours été appliquées pour les plasmides témoins et les fragments d'ADN utilisés dans les étapes de ligation. Un moyen de remettre à jour ce tableau serait de recalculer les concentrations en ADN des stocks de plasmides utilisés comme témoin de transformation par quantification sur gel, puis de recalculer l'ensemble des fréquences de transformation.

#### Influence de la toxicité du gène cloné sur la transformation normalisée

Les clonages n'entraînant pas de toxicité pour la souche hôte soit par l'absence de gènes toxiques dans l'insert ou dans le vecteur, soit par la présence d'un gène immunité dans la souche hôte, présentent une fréquence de transformation normalisée par mole d'ADN avoisinant  $10^0$  (**Tableau 5**). Dans les autres cas, cette fréquence est inférieure à  $10^{-2}$ . Le clonage de gènes toxiques est donc réalisable, mais à des fréquences d'au moins deux ordres de grandeur inférieures par rapport à d'autres gènes. Ces données mettent en exergue la difficulté de cloner dans un système d'expression inductible des gènes codant des fonctions toxiques pour la cellule, du fait d'un niveau rémanent d'expression du gène toxique.

Des clones *E. coli* ne présentant pas les constructions moléculaires escomptées (faux positifs) ont été obtenus avec une fréquence de transformation normalisée de l'ordre de  $10^{-2}$  suite aux problèmes techniques décrits ci-dessus. Après optimisation de chacune des étapes (digestion, déphosphorylation, transformation) pour diminuer le nombre de faux positifs, des échecs ont encore été rencontrés avec une fréquence de transformation normalisée TN par mole inférieure ou égale à  $10^{-5}$ . La probabilité d'obtention de la construction est dans ce cas toujours inférieure à la probabilité de recouvrement des plasmides matrices utilisés pour les amplifications ou des résidus de plasmides non digérés.



**Tableau 6 : Efficacité des constructions toxiques réalisées**

Plasmide	Souche <i>E. coli</i>	C (IPTG) ( $\mu\text{g.mL}^{-1}$ )	C (arabinose) ( $\text{g.L}^{-1}$ )	Nombre de reproduction	Efficacité moyenne	Ecart type
<b>pMR32</b>	DH5 $\alpha$	500		2	8,88E-06	1,22E-05
<b>pMRTC20</b>	DH5 $\alpha$ Z1	500		1	1,55E-05	
<b>pFEB98</b>	DH5 $\alpha$	500		3	1,82E-05	2,14E-05
<b>pB33C3</b>	TELI		10	2	1,72E-05	2,52E-06
<b>pZSCmC3</b>	DH5 $\alpha$ Z1		10	2	5,38E-05	1,63E-05

Ces résultats sont dans ce cas interprétés comme une trop forte toxicité des gènes clonés dans le système construit. Le système d'expression inductible choisi pour le gène toxique cloné ne serait pas adéquat. Le remplacement du système d'expression inductible toxique par d'autres systèmes plus stricts a effectivement permis l'obtention de transformants présentant la construction génétique attendue avec des fréquences TN par mole d'ADN supérieures à  $10^{-5}$  (**Tableau 5**).

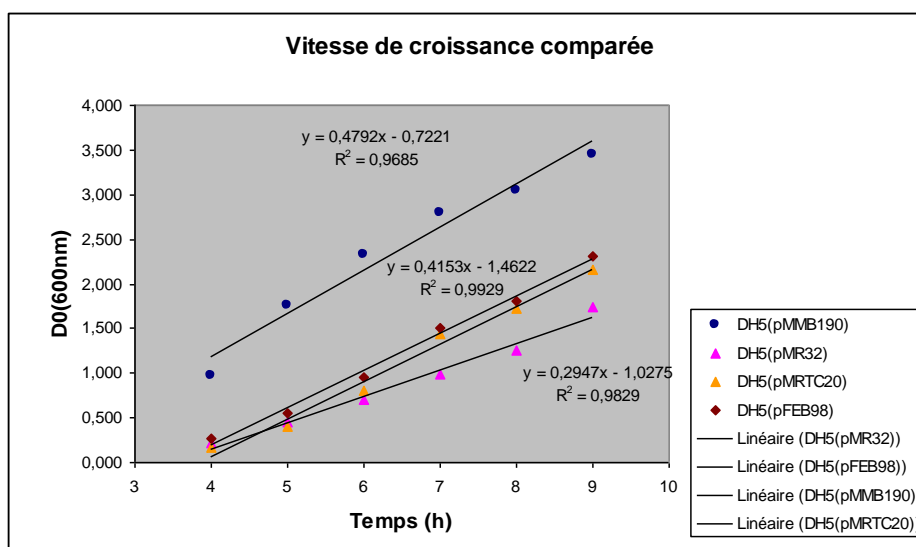
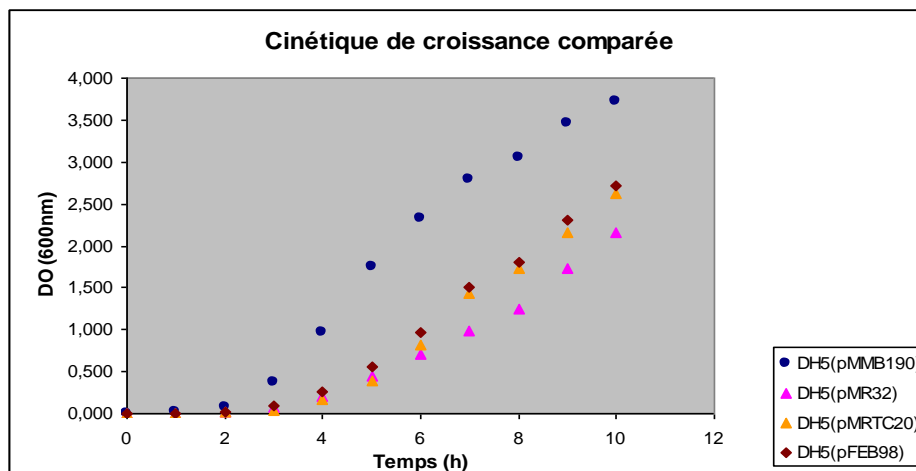
### 3.2.2 Tests de différents promoteurs pour le contrôle de l'expression du gène *relF*

#### Vérification de la fonctionnalité des constructions génétiques contenant le gène *relF*

Les plasmides pMR32 et pMRTc20, dérivés de pMMB190 dans lequel a été inséré le gène *relF* sous contrôle de promoteurs dérivés de  $P_{lac}$ , ont été construits ; ainsi que le plasmide pB33R, pour lequel le gène *relF* a été cloné dans le plasmide pBAD33 sous le contrôle du promoteur  $P_{araBAD}$  (également appelé  $P_{bad}$ ) (**Tableau 5**). Alors qu'une mortalité en condition inductrice des cellules hôtes contenant pMR32 et pMRTc20 est observée, les cellules hôtes contenant pB33R ne meurent pas en présence d'inducteur. L'expression de gènes dans pBAD33 sous le contrôle de  $P_{bad}$  a été vérifiée (voir résultats sur le gène *colE3*) et permet l'expression de protéines en condition d'induction. La quantité de protéines RelF exprimées par  $P_{bad}::relF$  en présence d'inducteur serait donc insuffisante pour entraîner la mort de son hôte. La quantité de protéines exprimées en condition inductrice est en effet plus faible si l'expression est contrôlée par  $P_{bad}$  comparativement à d'autres promoteurs dérivés  $P_{lac}$  (Guzman *et al.*, 1995; Newman & Fuqua, 1999). Aussi, un promoteur qui entraîne une forte expression en condition d'induction est indispensable pour une bonne efficacité de *relF* à tuer son hôte.

#### Efficacité des constructions fonctionnelles

Les constructions pMR32 et pMRTc20 ont une efficacité moyenne similaire avec un taux d'échappement de l'ordre de  $10^{-5}$  (**Tableau 6**). Nous avons cependant constaté pour DH5(pMR32) la formation d'un tapis bactérien après étalement de la dilution  $10^0$  de la culture bactérienne sur milieu gélosé contenant l'inducteur, et le développement de quelques colonies pour la dilution  $10^{-1}$ . Ce saut en UFC.mL<sup>-1</sup> de bactéries échappant à l'induction de l'expression toxique entre les dilutions  $10^0$  et  $10^{-1}$  a été confirmé par la reproduction de ces tests d'efficacité avec 500 µg.mL<sup>-1</sup> d'IPTG (2 répliquats), et avec 1 µg.mL<sup>-1</sup> d'IPTG (3



**Figure 7 : Cinétique de croissance et vitesse de croissance comparée de *E. coli* contenant les plasmides pMMB190, pMR32, pMRTC20 et pFEB98**

La densité optique ( $DO_{600nm}$ ) des cultures des différentes souches de *E. coli* a été mesurée à 600 nm toutes les heures, des étalements pour dénombrement ont été effectués toutes les 2 heures. Une cinétique de croissance similaire est obtenue dans les 2 cas. La vitesse de croissance de *E. coli* (pMMB190) étant supérieure à celle des autres souches, le nombre de mesure en UFC/mL pour cette *E. coli* (pMMB190) est insuffisante pour la comparaison des vitesses de croissance entre les différentes souches. Cette comparaison a donc été effectuée à partir des mesures de DO.

répliquats) et quelle que soit la densité cellulaire initiale ( $10^5$ ,  $10^8$  ou  $10^9$  UFC.mL<sup>-1</sup>). Le taux d'échappement de  $10^{-5}$  n'est obtenu qu'en tenant compte des dilutions inférieures à  $10^0$ , il serait sinon de seulement  $10^{-2}$ . Un tel phénomène n'a pas été observé pour le plasmide pMRTc20 pour des densités cellulaires similaires avant dilution.

#### Modification des constructions toxiques

De nombreux essais de clonage ont été réalisés sur pMR32. Les essais d'insertion dans pMR32 d'une cassette à bouts cohésifs de 3 kb, contenant à la fois le gène *relF* sous le contrôle de P<sub>A1-O3/O4</sub> et un gène de résistance à la tétracycline, ont tous échoué. Uniquement le gène de résistance *tc<sup>r</sup>* seul a pu être inséré dans pMR32. La présence du gène toxique *relF* dans la cassette expliquerait donc les échecs de clonage de [P<sub>A1-O3/O4</sub>::*relF*-*tc<sup>r</sup>*] dans pMR32. Cette même cassette a néanmoins pu être clonée en bouts francs dans le vecteur pMMB190, ne contenant pas de gène toxique. Les échecs de construction ne sont donc pas seulement liés à la cassette [P<sub>A1-O3/O4</sub>::*relF*-*tc<sup>r</sup>*] en elle-même, mais à la présence simultanée des deux cassettes [P<sub>taclacUV5</sub>::*relF*] et [P<sub>A1-O3/O4</sub>::*relF*-*tc<sup>r</sup>*] sur pMMB190. Un autre plasmide, pFEB98, contenant 2 gènes *relF* sous le contrôle de P<sub>A1-O3/O4</sub> a par contre pu être isolé (en une étape). La présence de deux gènes toxiques sur un même plasmide est donc techniquement réalisable. Cependant le clonage successif de deux gènes toxiques sur un même plasmide est délicat, et nécessite un parfait contrôle de l'expression par le système de régulation.

#### Cinétique de croissance et toxicité

Des cinétiques de croissance de *E. coli* contenant pMR32, pMRTc20, pFEB98 et pMMB190 (vecteur témoin ayant servi à la construction des trois premiers plasmides) ont été effectuées en milieu sélectif contenant de l'ampicilline uniquement. La vitesse de croissance des souches *E. coli* contenant des gènes toxiques dans leur plasmide est ralentie par rapport à *E. coli* contenant le plasmide témoin pMMB190. La vitesse de croissance de *E. coli* (pMR32) est 1,4 fois plus lente que celles de *E. coli* (pMRTc20) et *E. coli* (pFEB98) et 1,7 fois plus lente que celle de *E. coli* contenant pMMB190 (**Figure 7**). Ce ralentissement de croissance peut être interprété comme le résultat d'une expression toxique par la construction suicide (Bej *et al.*, 1988; Munthali *et al.*, 1996; Recorbet *et al.*, 1993). Bien qu'aucun inducteur n'ait été ajouté, une expression résiduelle du gène *relF* existerait. Le ralentissement de croissance étant plus important pour *E. coli* (pMR32) que pour *E. coli* (pMRTc20), l'expression basale des gènes *relF* serait plus importante avec pMR32. Le promoteur P<sub>taclacUV5</sub> serait donc beaucoup moins strict que P<sub>A1-O3/O4</sub> dans le contrôle de l'expression des gènes en absence

d'inducteur. Ces résultats sont en accord avec les données de la littérature, qui indiquent que  $P_{\text{taclacUV5}}$  est un promoteur hybride dont l'expression résiduelle (en absence d'inducteur) est supérieure à celle de  $P_{\text{lac}}$  (Lanzer & Bujard, 1988; Morales *et al.*, 1991), et donc de  $P_{\text{A1-O3/O4}}$  (**Annexe 7**). Malgré la présence de deux gènes *relF* dans pFEB98, le contrôle exercé via  $P_{\text{A1-O3/O4}}$  suffirait donc à réprimer l'expression résiduelle avec plus d'efficacité que  $P_{\text{taclacUV5}}$ .

#### Stabilité des constructions toxiques

Après 10 h de cinétique de croissance, 95 colonies bactériennes étalées sur milieu solide pour le dénombrement d'UFC.mL<sup>-1</sup> ont été inoculées en microplaque puis testées pour leur sensibilité à l'expression des gènes *relF* par induction en milieu liquide. Plus de 90 % des souches *E. coli* (pMR32) sont résistantes à l'induction de l'expression du système toxique, contre 0 à 5 % seulement pour *E. coli* (pFEB98). Ainsi, des souches *E. coli* (pMR32) ayant acquis la capacité à échapper à l'induction du système suicide auraient été sélectionnées au cours de la croissance de la souche. La trop forte expression résiduelle de *relF* entraîne donc des problèmes de stabilité dans le temps d'un système toxique inductible fonctionnel. L'augmentation de l'activité résiduelle de l'expression toxique entraîne ainsi une augmentation du nombre de cellules devenues résistantes à l'expression toxique. Ces résultats rejoignent l'hypothèse de réponse hormétique des bactéries pour différentes concentrations d'une molécule toxique : une quantité insuffisante de molécules toxiques entraîne l'effet inverse de la réponse attendue (Calabrese, 2004). Dans le cas présent, la forte expression résiduelle de la protéine toxique pour la cellule en quantité insuffisante (forte expression basale) entraîne la sélection de cellules devenues résistantes à l'expression de *relF*. Aussi, le contrôle de l'expression de *relF* pour son application comme gène toxique inductible nécessite une expression résiduelle la plus faible possible pour que ce dernier maintienne une efficacité tueuse.

### 3.2.3 Développement d'un nouveau système de régulation pour le contrôle de l'expression du gène *colE3*

#### Essais de construction toxique contenant le gène *colE3*

Le clonage du gène *colE3* a dans un premier temps été testé sous le contrôle d'un promoteur de type  $P_{\text{lac}}$  strict, le promoteur  $P_{\text{A1-O3/O4}}$ . Le fragment d'ADN de pMMS3 contenant à la fois *colE3* sous contrôle de  $P_{\text{A1-O3/O4}}$  et un gène de résistance au chloramphénicol *cm<sup>r</sup>* a été introduit dans pMMB190 pour former le plasmide pMC. Le

fragment d'ADN cloné ne semble cependant pas exprimer de fonction toxique, comme l'indiquent premièrement la fréquence élevée de transformation de pMMS3 dans une souche *E. coli* S17.1  $\lambda$ pir (capable de répliquer ce plasmide) alors qu'aucun gène répresseur *lacI* n'est présent ni sur le vecteur ni dans la souche hôte, et deuxièmement la fréquence de transformation normalisée anormalement élevée ( $\text{TN.mol}^{-1} > 0$ ) de pMC dans une souche *E. coli* non immune (**Tableau 5**). Ces résultats préliminaires ont été confirmés par des tests d'induction sur *E. coli* EC100 (pMC) qui se sont révélés négatifs. Le gène *colE3* s'il est présent n'est donc pas fonctionnel. Les profils de restriction de pMC confirment cependant la présence d'un insert de la taille attendue, mais le séquençage des inserts clonés indique de multiples remaniements génétiques entre le gène *colE3* et le promoteur  $P_{A1-O3/O4}$  (**Annexe 7**). Il semblerait que des événements mutationnels se soient produits, mais nous ne savons pas si ces mutations sont présentes dans le plasmide initial pMMS3 ou s'ils ont eu lieu lors des manipulations génétiques. Nous sommes actuellement en attente des séquences de la cassette contenant *colE3* de pMMS3 qui permettront de répondre à ces questions.

Le gène *colE3* a également été cloné sous le contrôle du promoteur plus strict  $P_{\text{bad}}$  de pBAD33, mais aucune construction toxique n'a pu être recouvrée dans *E. coli* JM109. Un examen de la littérature révèle que ce gène, qui a été exploité dans l'endiguement génétique (Diaz *et al.*, 1994) et comme système de contre sélection positive pour le clonage de gènes (Vernet *et al.*, 1985), n'a été stabilisé que dans des souches présentant un gène immunité. Afin de vérifier l'origine technique ou biologique des échecs de clonage de *colE3* sous le contrôle de  $P_{\text{bad}}$ , les produits de ligation pBAD33 + *colE3* ont été transformés dans une souche produisant constitutivement l'antitoxine de ColE3, la protéine ImmE3 (souche *immE3+*) (Jakes *et al.*, 1974). Des clones comportant la construction plasmidique pB33C3 attendue ont pu être isolés, mais la réintroduction de pB33C3 dans une souche non immune (*immE3-*) n'a pas permis la croissance de transformants. Les difficultés rencontrées pour cloner *colE3* ne sont donc pas liées à un problème technique, mais à la toxicité exprimée par le gène cloné. La très faible expression résiduelle existant via  $P_{\text{bad}}$  serait suffisante pour entraîner la mort de la souche hôte, empêchant ainsi le recouvrement de la construction toxique contenant *colE3* cloné dans pBAD33. Un tel phénomène a également été observé pour d'autres gènes exprimant des fonctions hautement toxiques pour la cellule (Torres *et al.*, 2000).

Les connaissances apportées par l'étude approfondie de différents promoteurs dans *E. coli* ont permis l'identification des facteurs intervenant dans l'efficacité des promoteurs. L'affinité de l'ARN polymérase pour la séquence promotrice et l'emplacement des opérateurs, peuvent avoir un impact sur l'expression résiduelle en absence d'inducteur

(Lanzer & Bujard, 1988). Lutz *et al.* ont construit une série de promoteurs hybrides, résultant de la combinaison de promoteurs et opérateurs provenant de différents systèmes régulateurs, pour à la fois brider l'expression basale et favoriser une expression maximale en condition inductrice (Lutz & Bujard, 1997). L'un de ces promoteurs  $P_{lac/ara1}$ , présent sur un vecteur à très faible nombre de copies (4 copies par cellule), a été testé pour cloner *colE3* en absence d'un gène immunité. Ce promoteur permettrait d'après les auteurs un contrôle strict de l'expression en condition non inductrice par la transcription de seulement un ARNm toutes les 3 générations (Lutz & Bujard, 1997). Le gène *colE3* a donc été cloné sous le contrôle de  $P_{lac/ara1}$  dans pZSCm, plasmide dérivé de pZS24MCS1 (Lutz & Bujard, 1997), modifié par remplacement du gène de résistance à la kanamycine par le gène de résistance au chloramphénicol. Après transformation du produit de ligation pZSCm+*colE3* dans la souche DH5 $\alpha$ Z1 (*immE3*-), des transformants possédant le plasmide pZSCmC3 ont pu être isolés. Le gène *colE3* porté par pZSCmC3 est parfaitement fonctionnel puisque l'ajout d'inducteurs entraîne la mort des cellules hôtes. Ces résultats montrent qu'il est donc possible de cloner le gène *colE3* en absence de son gène immunité, confirmant d'autres travaux publiés en 2004 par Anthony *et al.* puis par Bowers *et al.* (2004). Ces derniers ont également réussi à cloner *colE3* sous le contrôle du promoteur  $P_{lac}$  en diminuant le nombre de copies du vecteur et en ajoutant des terminateurs en amont du gène cloné (pour éviter toute transcription par des promoteurs cryptiques ou par des gènes orientés dans le même sens que le gène cloné) (Anthony *et al.*, 2004; Bowers *et al.*, 2004).

Il est cependant intéressant de préciser que les colonies de DH5 $\alpha$ Z1 (pZSCmC3) obtenues ont un aspect translucide et sont de taille inférieure aux colonies de la souche DH5 $\alpha$ Z1. Bien que des cinétiques de croissance n'aient pas encore été conduites sur ces souches, un retard de croissance de DH5 $\alpha$ Z1 (pZSCmC3) par rapport à DH5 $\alpha$ Z1 ou DH5 $\alpha$ Z1 (pZSCm) a été constaté. La stabilité dans le temps de la construction pZSCmC3 n'a pas été testée. Il serait intéressant de vérifier si la présence de *ColE3* entraîne comme pour *RelF* une réponse hormétique de la cellule hôte. La sélection de mutants pour *relF* a lieu pour une forte expression basale de la protéine *RelF* (supérieur à 1500 Unités  $\beta$ -galactosidase) (Lanzer & Bujard, 1988; Morales *et al.*, 1991). Si réponse hormétique de *ColE3* il y a, celle-ci est entraînée pour des quantités de protéines synthétisées inférieures à 4 Unités  $\beta$ -galactosidase (Lanzer & Bujard, 1988). A l'examen des résultats présentés pour pMR32, il nous a semblé préférable de développer un nouveau système de régulation pour le contrôle de l'expression inductible de *colE3*, qui préviendrait toute expression basale du gène *colE3*.

Développement d'un système hôte-vecteur pour le contrôle de l'expression de *colE3*

Outre les précautions usuelles quant au nombre de copies de vecteurs et l'ajout de terminateurs en amont du gène toxique cloné (Anthony *et al.*, 2004; Bowers *et al.*, 2004), d'autres stratégies pour le clonage de gènes hautement toxiques ont été développées, afin de limiter leur expression basale en condition non inductrice (Saida *et al.*, 2006). Un promoteur peut être placé en sens inverse du gène sous contrôle inductible. L'ARNm antisens alors synthétisé s'hybride avec l'ARNm sens produit en cas d'expression basale (Szafranski *et al.*, 1997). Des codons stops ambres peuvent également être clonés en amont du gène toxique. Ce codon stop empêche la traduction dans les souches sup0 (sans activité suppression ambre). La transcription n'est assurée que dans les seules souches sup+, reconnaissant le codon ambre comme codon initiateur (Ohashi-Kunihiro *et al.*, 2006; Saida *et al.*, 2006).

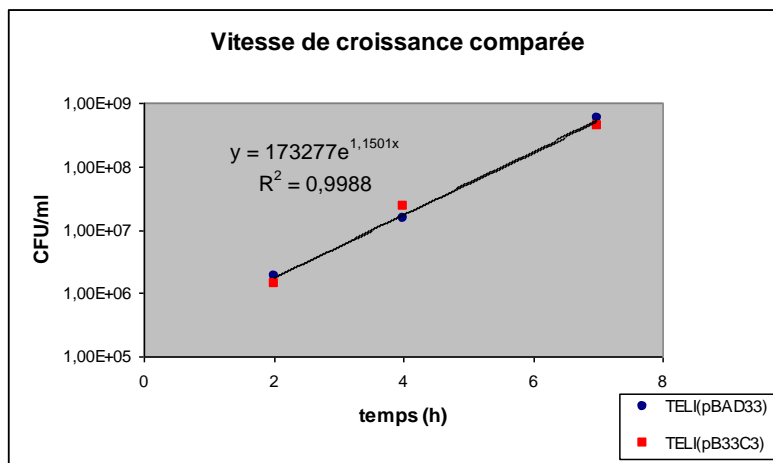
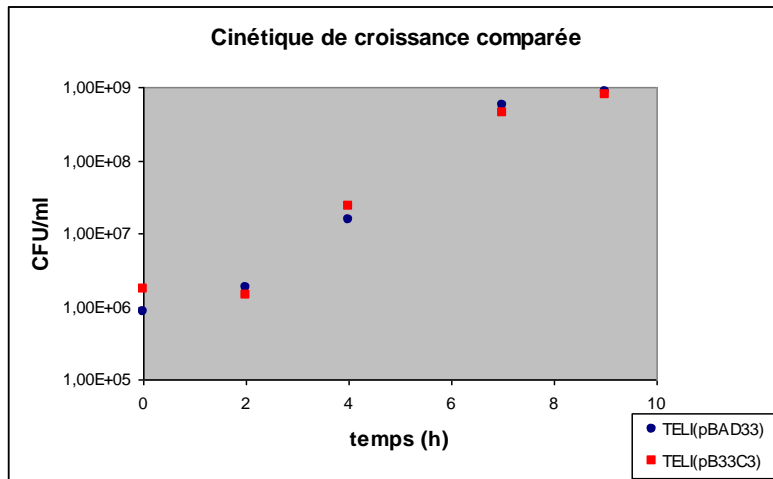
Un nouveau système de contrôle de l'expression des gènes *colE3* a été développé dans cette étude, basée sur la modification du ratio entre colicines et antitoxines produites en condition inductrice. La protéine ImmE3 interagit avec la protéine ColE3 selon un ratio 1 :1, inhibant ainsi l'activité toxique de la protéine (Jakes & Zinder, 1974). L'expression résiduelle de *colE3* peut donc être contrebalancée par la production continue (constitutive) de l'antitoxine ImmE3, permettant ainsi la stabilisation du gène *colE3*. Le succès de cette stratégie de contrôle repose sur la surexpression de *colE3* par rapport à *immE3* en condition inductrice, la quantité de colicines produites devant dépasser celle de ImmE3 pour entraîner la mort de la cellule hôte. Cette surexpression de *colE3* en condition inductrice a par exemple été évaluée chez la souche *E. coli* CC118 *immE3* (pEDF5). Le plasmide pEDF5 porte un gène *colE3* sous le contrôle d'un promoteur de type P<sub>lac</sub>, dont la fonctionnalité a été vérifiée au préalable. L'ajout d'IPTG n'entraîne cependant pas la mort de *E. coli* CC118 *immE3* (pEDF5), indiquant l'expression en condition inductrice d'une quantité de colicine inférieure à celle de l'antitoxine ImmE3. La quantité de protéine ImmE3 exprimée constitutivement dans cette souche est beaucoup trop élevée pour permettre la mort de la cellule hôte en condition inductrice. La construction d'une souche exprimant constitutivement *immE3* mais à faible taux est nécessaire au succès de cette stratégie de contrôle. Une nouvelle souche *E. coli* *immE3+* a donc été construite.

Des amorces ont été définies sur pColCA38 (Masaki & Ohta, 1982) pour amplifier le gène *immE3*. Ce plasmide contient en effet les gènes *colE3* et *immE3* dans le même opéron. Les deux séquences codantes étant séparées par seulement 9 nucléotides, un soin particulier a été apporté à la définition des amorces pour amplifier uniquement la séquence codante (RBS



ou site de fixation du ribosome compris) du gène *immE3*, sans élément de *colE3*. Ce gène a été cloné sous le contrôle d'un promoteur du gène de résistance à l'ampicilline  $P_{bla}$ , promoteur connu pour sa faible expression constitutive. Puis la cassette  $P_{bla}::immE3$  a été transférée dans le chromosome d'une souche *E. coli* TE2680 (Elliott, 1992) grâce au vecteur pRS551 (Simons *et al.*, 1987) permettant l'intégration de la cassette dans l'opéron *trp* génomique de la souche par double recombinaison. Dans cette cassette, le gène répresseur *lacI<sup>f</sup>* a été ajouté en perspective d'utilisation de la souche modifiée avec les systèmes toxiques décrits ultérieurement. L'intégration est contrôlée grâce au gène de résistance à la kanamycine (*kn<sup>r</sup>*) co-transféré avec la cassette, et à la perte du gène de résistance au chloramphénicol (*cm<sup>r</sup>*) suite à la double recombinaison. Pour la réalisation d'une telle cassette, une partie des séquences génétiques de pRS551 servant à la recombinaison / intégration dans *E. coli* TE2680 (4 kb sur 5 kb) a dû être éliminée. Le vecteur pRSLI1 construit a été linéarisé puis intégré dans TE2680. Les transformants ont été sélectionnés sur milieu de culture contenant de la kanamycine. Cependant, les transformants obtenus ne présentaient pas le phénotype attendu (conservation de la résistance au chloramphénicol), ce qui nous a contraint à redéfinir le vecteur d'intégration pRS551 + *ImmE3* + *lacI<sup>f</sup>*, en éliminant seulement 1,2 des 5 kb de la région d'ADN dévolue à la recombinaison. Le vecteur pRSLI2 construit a subi le même traitement que pRSLI1 avant transformation dans *E. coli* TE2680. Dans ces conditions, des transformants ayant le phénotype attendu ont ainsi pu être isolés.

La souche *E. coli* TELI alors construite est capable de se développer sur milieu gélosé contenant de la colicine E3, contrairement à la souche *E. coli* TE2680, confirmant l'expression du gène *immE3* dans la souche construite. Après transformation de pB33C3 dans *E. coli* TELI des transformants contenant le plasmide ont été obtenus. Aucune modification de taille des colonies de *E. coli* TELI (pB33C3) n'a été constatée et les colonies n'ont plus l'aspect translucide observé précédemment. La quantité de protéine ImmE3 produite est donc suffisante pour permettre la stabilisation du plasmide contenant le gène *colE3*. En présence de l'inducteur, *E. coli* TELI (pB33C3) est tué. Le système régulateur hôte-vecteur construit est donc fonctionnel, avec un taux d'échappement de  $10^{-5}$  (**Tableau 6**). Contrairement au précédent système testé, le système régulateur d'expression inductible développé est donc parfaitement fonctionnel, sans les répercussions des effets toxiques visuellement constatés par l'aspect des colonies.



**Figure 8 : Cinétique de croissance et vitesse de croissance comparée de souches *E. coli* TELI contenant les plasmides pBAD33 et pB33C3**

La densité optique (DO<sub>600nm</sub>) des cultures des différentes souches de *E. coli* a été mesurée à 600 nm toutes les heures, des étalements pour dénombrement sont effectués toutes les 2 heures.

### Cinétique de croissance et toxicité résiduelle

Des cinétiques de croissance de *E. coli* TELI contenant les plasmides pBAD33 et pB33C3 ont été conduites en milieu sélectif contenant du chloramphénicol. La vitesse de croissance des deux souches est similaire, aucun retard de croissance n'étant constaté (**Figure 8**). La présence du gène immunité dans le chromosome de la souche hôte évite donc l'expression d'une toxicité résiduelle, favorisant ainsi la stabilité de la construction toxique. Contrairement à tous les systèmes précédemment testés, le nouveau système de contrôle de l'expression inductible développé permet donc la mort des cellules hôtes en condition inductrice avec un faible taux d'échappement, tout en évitant les effets nocifs sur la cellule hôte en condition non inductrice, effets liés à l'expression résiduelle.

## **4 CONCLUSION**

Du fait de leur expression toxique, la manipulation des gènes toxiques est très délicate et nécessite une optimisation des techniques de biologie moléculaire pour limiter tout bruit de fond couvrant la détection des transformants possédant la construction toxique, ainsi que de nombreux contrôles afin de vérifier l'échec ou la réussite des clonages. A ce titre, l'évaluation de la fréquence de transformation des produits de ligation (par rapport aux contrôles) a été un critère que nous avons utilisé pour juger l'échec ou la réussite du recouvrement des constructions toxiques. Notamment, une fréquence de transformation normalisée par mole d'ADN inférieure à  $10^{-5}$  (selon les méthodes de quantification utilisées) est généralement indicatrice d'un échec de clonage, lié à un système de contrôle de l'expression inductible non adéquate. La valeur de ce critère et la définition du seuil nécessiteraient cependant la répétition des expérimentations et le réajustement des méthodes de quantification entre le plasmide témoin et les fragments d'ADN utilisés dans l'étape de ligation.

Avec les méthodes développées, nous avons néanmoins réussi à cloner deux gènes toxiques sous différents promoteurs, permettant de choisir le système régulateur le plus adéquat pour le contrôle de leur expression. Les résultats obtenus montrent l'impact que l'expression résiduelle des gènes toxiques en condition non inductrice peut avoir sur le clonage des gènes toxiques, ainsi que sur la stabilité des constructions toxiques. Notamment, une trop forte expression résiduelle du gène *relF* entraîne la sélection de clones résistants à l'expression de *relF* en condition inductrice. Le contrôle de l'expression basale est donc une étape primordiale vers la construction de systèmes toxiques efficaces. Un point également

important à considérer est la quantité de protéines exprimées en présence d'inducteur, l'activité létale liée à *relF* nécessitant une forte expression de la protéine en condition inductrice, alors que celle liée à *colE3* est obtenue pour des niveaux d'expression très faibles.

Le choix d'un système régulateur doit ainsi être adapté en fonction du gène toxique utilisé. Ainsi le gène *colE3* nécessite une expression basale quasiment nulle pour être cloné. Une très faible quantité de colicine est en effet suffisante pour entraîner la mort de la cellule hôte. Le gène *relF* est plus permissif à une expression basale qui doit cependant rester suffisamment faible pour éviter la sélection de clones résistants et une forte expression en condition inductrice pour tuer l'hôte. Le promoteur  $P_{A1-O3/O4}$  est celui ayant donné les meilleurs résultats pour assurer l'expression inductible du gène *relF*. Pour le contrôle de l'expression inductible de *colE3*, nous avons développé un nouveau système régulateur permettant la stabilisation du gène *colE3* dans l'hôte en absence d'inducteur et une activité létale efficace en condition inductrice.

## **SOUS CHAPITRE III-2**

### **DEVELOPPEMENT D'UN OUTIL DE CONTRE SELECTION POSITIVE POUR PIEGER L'ADN DE LA MICROFLORE TELLURIQUE**

#### **1 INTRODUCTION**

La nouvelle approche métagénomique proposée dans cette thèse repose sur l'utilisation d'une souche bactérienne réceptrice génétiquement préparée pour piéger des gènes de bactéries cultivées, comme des non cultivées. Cette stratégie est basée sur la présence de régions d'ADN, sur lesquelles peut s'initier la recombinaison de l'ADN du métagénome. Un double événement de recombinaison permet l'intégration de l'ADN exogène présentant une similarité de séquence suffisante avec les régions cibles définies, et entraîne l'élimination par échange allélique d'une cassette toxique présente entre les deux régions d'ADN cibles. Les cellules recombinées peuvent alors être spécifiquement isolées, l'induction des gènes létaux de la cassette toxique conduisant à la mort des autres cellules. L'intérêt de cette approche repose entre autres sur l'obtention directe des souches recombinantes ayant capturé l'information génétique ciblée de la microflore tellurique, sans autre étape de criblage.

Les travaux précédents sur la définition de la cassette toxique ont montré l'importance du choix du système régulateur contrôlant son expression. Lorsque le système régulateur utilisé n'est pas adéquat, des clones résistants à l'expression toxique peuvent être sélectionnés (sous chapitre III-1), pouvant ainsi affecter l'efficacité globale du système de contre sélection positive à sélectionner les seules souches réceptrices. Un autre point à considérer dans le développement de cet outil de capture est l'aptitude des souches réceptrices à recombiner des séquences d'ADN plus ou moins divergentes. L'augmentation de la divergence des séquences entraîne en effet une diminution de la fréquence de recombinaison (Vulic *et al.*, 1999). Dans notre approche, où la capture d'une diversité de séquences est visée, l'emploi d'une souche réceptrice hautement recombinogène et / ou acceptant la recombinaison de séquences divergentes par rapport aux régions d'ADN cibles du système de contre sélection permettrait d'élargir le crible des informations génétiques recherchées.

La recombinaison de l'ADN dans le génome bactérien a été particulièrement étudiée chez *E. coli*. La recombinaison homologue fait intervenir le complexe enzymatique RecBCD et la protéine RecA (Dabert & Smith, 1997). La protéine RecA est notamment l'enzyme clé catalysant la réaction de recombinaison entre l'ADN simple brin et l'ADN génomique (Chaudhury & Smith, 1984). L'ADN double brin entrant dans la cellule est dégradé par le complexe enzymatique RecBCD jusqu'à la rencontre de séquences *chi* (5' GCTGGTGG 3'). La sous unité RecD perd alors son affinité pour l'hétérodimère RecBC, qui devient déficient en activité exonucléasique mais proficient en activité hélicasique. Le complexe RecBC dissocie alors les deux brins d'ADN et la protéine RecA s'associe alors à l'ADN simple brin nouvellement formé. L'intégration est ensuite amorcée si une homologie de séquences suffisante est détectée (Matic *et al.*, 1996). Une séquence minimale d'identité (MEPS ou *minimal efficient processing segment*) est en effet requise (Shen & Huang, 1986). Des régions de forte similarité de 25 à 200 pb suivant les espèces sont nécessaires pour initier la complémentation entre l'ADN entrant et l'ADN de la souche réceptrice (Thomas & Nielsen, 2005). La fréquence de recombinaison augmentant avec la taille de la zone homologue, des régions de similarité de 1 kb sont souvent utilisées pour permettre l'intégration dans le génome d'une souche réceptrice d'un fragment de plusieurs kilobases (de Vries & Wackernagel, 2002).

Un outil de plus en plus utilisé pour activer la recombinaison est le système lambda red gam. Ce système est constitué de gènes du phage lambda, exprimés pour permettre sa réplication et son intégration dans le génome de la souche hôte (Poteete *et al.*, 1988). Les gènes *exo* et *bet* (constituant le système red) du bactériophage permettent la recombinaison et l'intégration de l'ADN dans la souche hôte. Le gène *exo* exprime une 5'-3' exonucléase. Le gène *bet* exprime une protéine se fixant sur l'ADN simple brin, stimulant l'hybridation avec l'ADN génomique. L'invasion de l'ADN simple brin nécessite cependant l'action de la protéine RecA (Delmas & Matic, 2005; Muniyappa *et al.*, 1984) ainsi que les protéines Gam (Murphy, 1998). Ces dernières se fixent sur le complexe enzymatique RecBCD, inhibant ainsi entre autres l'activité exonucléasique du complexe (Murphy, 1991). Chez *E. coli*, ces systèmes augmentent les fréquences de recombinaison d'ADN linéaire par rapport à la voie RecBCD (Murphy, 1998) et permettent l'amorçage de la recombinaison avec des séquences de seulement 30-50 nucléotides (Copeland *et al.*, 2001; Datsenko & Wanner, 2000; Poteete, 2001; Yu *et al.*, 2000).

**Tableau 7 : Liste des souches *E. coli* et plasmides utilisés ou construits**

Souche <i>E.coli</i>	Génotype	Phénotype	Référence
<b>TELI</b>	F- $\lambda$ - IN(rrnD-rrnE)1 $\Delta$ (lac)X74 rpsL galK2 recD1903::Tn10d-Tet trpDC700::putA1303::[kn <sup>r</sup> -lacI <sup>q</sup> -immE3-lac]	Kn <sup>r</sup> , Tc <sup>r</sup> , ImmE3 <sup>+</sup> , Cm <sup>s</sup>	Cette étude
<b>1661<sup>a</sup></b>	F- lac-3350, galK2, galT22, recD1014, IN(rrnD-rrnE)1, rpsL179	Sm <sup>r</sup> , ImmE3 <sup>+</sup>	Chaudhury, 1984
<b>1661LI</b>	F- lac-3350, galK2, galT22, recD1014, IN(rrnD-rrnE)1, rpsL179 $\Delta$ trp::[Kn <sup>r</sup> -immE3-lacI <sup>q</sup> ]	Sm <sup>r</sup> , Kn <sup>r</sup> , Trp <sup>-</sup> , ImmE3 <sup>+</sup>	Cette étude
<b>1661LIAara</b>	F- lac-3350, galK2, galT22, recD1014, IN(rrnD-rrnE)1, rpsL179 $\Delta$ trp::[Kn <sup>r</sup> -immE3-lacI <sup>q</sup> ] $\Delta$ araBAD	Sm <sup>r</sup> , Kn <sup>r</sup> , Trp <sup>-</sup> , Ara <sup>-</sup> , ImmE3 <sup>+</sup>	Cette étude
<b>BW25141</b>	lacI <sup>q</sup> rrnB <sub>T14</sub> $\Delta$ lacZ <sub>WJ16</sub> $\Delta$ phoBR580 hsdR514 $\Delta$ araBAD <sub>AH33</sub> $\Delta$ rhaBAD <sub>LD78</sub> galU95 endA <sub>BT333</sub> uidA( $\Delta$ mluI)::pir <sup>+</sup> recA1	Ara <sup>-</sup>	Datsenko <i>et al.</i> , 2000

Plasmide	Génotype	Phénotype	Référence
<b>pBAD33</b>	cm <sup>r</sup> , p15A, araC, P <sub>bad</sub>	Cm <sup>r</sup>	Guzman <i>et al.</i> , 1995
<b>pB33C3</b>	cm <sup>r</sup> , p15A, araC-P <sub>bad</sub> ::colE3	Cm <sup>r</sup> , Ara <sup>s</sup>	Cette étude
<b>pPR13</b>	cm <sup>r</sup> , p15A, PheR1-PheR2	Cm <sup>r</sup>	Cette étude
<b>pPRE</b>	cm <sup>r</sup> , p15A, PheR1-araC -P <sub>bad</sub> ::colE3-PheR2	Cm <sup>r</sup> , Ara <sup>s</sup>	Cette étude

<sup>a</sup> Souche *E. coli* V355, Coli Center Genetic Stock Center (CGSC) 6720

Ara<sup>-</sup> : auxotrophe pour l'arabinose, Ara<sup>s</sup> : sensible à l'induction arabinose, Cm<sup>r</sup> : résistant au chloramphénicol, Cm<sup>s</sup> : sensible au chloramphénicol, ImmE3<sup>+</sup> : production d'antitoxine ImmE3, Kn<sup>r</sup> : résistant à la kanamycine, Sm<sup>r</sup> : résistant à la streptomycine, Tc<sup>r</sup> : résistant à la tétracycline, Trp<sup>-</sup> : auxotrophe pour le tryptophane

Les gènes *exo*, *bet*, et *gam*, provenant du phage lamda, ont été clonés sur des vecteurs pour leur emploi en biologie moléculaire (Datsenko & Wanner, 2000). Ce système red gam est généralement placé sous le contrôle de promoteurs induits par l'IPTG ou l'arabinose pour activer de manière contrôlée la recombinaison (Chaveroche *et al.*, 2000; Datsenko & Wanner, 2000; Murphy, 1998). Une série de plasmides récemment développés permet l'expression du système red gam par une augmentation de la température à 42°C (Datta *et al.*, 2006). Les gènes sont sous le contrôle du promoteur P<sub>L</sub>, dont l'expression est réprimée par la protéine thermosensible (ts) CI857. La dénaturation de la molécule à 42°C permet ainsi une dépression du promoteur P<sub>L</sub> et donc l'expression des gènes du système red gam. Les systèmes lamda red gam sont efficaces chez *E. coli* (Murphy, 1998). Leur fonctionnement a également été rapporté chez *Salmonella enterica* (Datta *et al.*, 2006)

Pour valider le concept de piégeage d'une information génétique ciblée de la microflore tellurique dans une souche réceptrice, nous avons développé un système de contre sélection positive dans *E. coli*. Cette souche facilement manipulable génétiquement autorise également l'utilisation de systèmes lambda red gam pour activer la recombinaison. Des premiers tests expérimentaux de transformation d'ADN métagénomique dans la souche réceptrice ont été effectués *in vitro*. L'objectif de ces essais était de vérifier, d'une part la capacité du système à sélectionner les souches recombinées, d'autre part l'efficacité de la recombinaison entre l'ADN métagénomique et les régions ADN cibles introduit dans la souche réceptrice. Les développements effectués et les résultats obtenus sont ici présentés.

## **2 MATRIELS ET METHODES**

### **2.1 Bactéries et plasmides**

Les souches de *E. coli* et les plasmides construits ou utilisés sont répertoriés dans le **Tableau 7**. Le stockage et la culture bactérienne ont été effectués comme décrit dans le sous chapitre III-I, en présence des antibiotiques adéquats selon les concentrations précisées dans le sous chapitre III-1. Du L(+) arabinose (Carl Roth, Lauterbourg, France) à 10 g.L<sup>-1</sup> a été ajouté dans le milieu quand précisé. Les incubations ont été effectuées à 37°C pour la majorité des souches *E. coli*. Celles contenant pSIM6 ont été manipulées à 30°C sauf lorsque précisé. Les cultures liquides ont été effectuées avec agitation orbitale à 150 rpm sauf lorsque précisé.



**Tableau 8 : Liste des amorces, de leurs caractéristiques et des conditions d'amplification utilisées**

Amorce	Séquence en 5'-	ADN cible	Site restriction	Gène amplifié	Mésappariements	Tm <sup>a</sup> (°C)
<b>Bad33bglIIIFw2</b>	TTGATTACGCGCCCTGCAGC	pBAD33	PstI	Ori p15A, <i>cm<sup>r</sup></i>	1	63
<b>Bad33pstIRv2</b>	CACCGATGGGGAAGATCTGGC	pBAD33	BglII	Ori p15A, <i>cm<sup>r</sup></i>	1	63
<b>PbcolE3notIFw5</b>	GATCTGCTCATGTGCGGCCGCTTATCA	pB33C3	NotI	<i>araC-P<sub>bad</sub>::colE3</i>	4	59
<b>PbcolE3ndeIRv5</b>	CAAAGCATATGTAGAAACGCAAAAGGC	pB33C3	NdeI	<i>araC-P<sub>bad</sub>::colE3</i>	3	59
<b>PheR1f</b>	GACAAGATCTAGATCTAGATCTGTAGCCGCGCCAGAACCATTTGT	métagénome	BglII	<i>phe</i>	0	55
<b>PheR1r</b>	ACTAATGCATATGCATGCGGCCGCGATCATCGACGCGTTCCAGCAGA	métagénome	AvaIII <sup>d</sup> , NdeI, NotI	<i>phe</i>	0	55
<b>PheR2f</b>	ACTAATGCATATGCATGCGCGCGCCGCGATAGATCAGACCATCG	métagénome	AvaIII <sup>d</sup> , NdeI, SgsI <sup>e</sup>	<i>phe</i>	0	50
<b>PheR2r</b>	GACACTGCAGCTGCAGGACTGGTATGCGCTGAAGGA	métagénome	PstI	<i>phe</i>	0	50
<b>Bad342Fw1</b>	TTTAAACGTGGCCAATATGGAC	pPR13, pPRE		<i>pheR1-X-pheR2<sup>c</sup></i>	0	55
<b>Bad342Rv1</b>	AGTTCTGCTATCAATGGCCTAT	pPR13, pPRE		<i>pheR1-X-pheR2<sup>c</sup></i>	0	55

Les sites de restriction sont indiqués sur la séquence par un soulignement

<sup>a</sup> Tm : température d'hybridation a été définie expérimentalement

<sup>b</sup> Ajout de mésappariements par rapport à la séquence nucléotidique de l'ADN cible

<sup>c</sup> Amplification de la séquence nucléotidique comprenant pheR1, pheR2, et les séquences présentes entre les deux sites de recombinaison

<sup>d</sup> AvaIII=NsiI,

<sup>e</sup> SgsI= AscI

Les plasmides ont été extraits avec le kit Macherey Nagel NucleoBond PC100 et NucleoSpin Plasmid (Hoerd, France) selon les recommandations du fournisseur. La qualité et la concentration de l'extrait plasmidique ont été vérifiées par dosage au spectrophotomètre ND-1000 (NanoDrop Technologies, Wilmington, United States of America).

## 2.2 Manipulations enzymatiques de l'ADN

Les réactions d'amplification ont été réalisées comme décrit dans le sous chapitre III-1. Les amorces ont été définies sous oligo 6.1 et commandées chez InvitroGen (Cergy Pontoise, France). Les amorces et les conditions d'amplification utilisées sont répertoriées dans le **Tableau 8**. Les réactions de digestion ont été effectuées avec les enzymes de restrictions adéquates (Fermentas, Mundolsheim, France) selon les recommandations du fournisseur. Les fragments d'ADN ont été traités lorsque précisé avec l'enzyme Klenow Fragments (Roche Diagnostics, Meylan, France) et l'enzyme Phosphatase Alcaline Shrimp SAP (Roche Diagnostics) selon les recommandations du fabricant. La taille des fragments d'ADN a été contrôlée par migration sur gel d'agarose en présence de marqueur de taille (1Kb+, InvitroGen, Cergy Pontoise, France). Les fragments d'ADN ont été purifiés à l'aide du kit Macherey Nagel NucleoSpin Extract II (Hoerd, France), soit directement sur colonne, soit à partir d'un gel (**Annexe 3**), puis quantifiés visuellement sur gel en présence de marqueur de masse (MassRuler, Fermentas, Mundolsheim, France). Les fragments ADN ont été ligués avec l'enzyme T4 DNA ligase (Roche Diagnostics) selon les recommandations du fabricant.

## 2.3 Construction du plasmide pPRE

Le vecteur pB342 a été amplifié par le couple d'amorces Bad33BglIIIFw2 / Bad33PstIRv2. Le produit PCR a été purifié sur colonne, digéré par PstI, BglII, puis déphosphorylé. Le site de recombinaison pheR1 a été amplifié par le couple d'amorces PheR1f / PheR1r, le produit PCR a été purifié sur colonne, traité avec l'enzyme Klenow Fragments, puis digéré par BglII. Le site de recombinaison pheR2 a été amplifié par le couple d'amorce PheR2f / PheR2r purifié sur gel, traité avec l'enzyme Klenow Fragments, puis digéré par PstI. Ces trois fragments ADN ont été purifiés puis ligués selon le ratio vecteur :R1 :R2 de 1 :5 :5, avec l'enzyme T4 ligase 1 nuit à 16°C. Le produit de ligation a été transformé dans *E. coli* DH5α chimiocompétentes (InvitroGen, Cergy Pontoise, France) selon le protocole fourni par le fabricant. Les transformants ont été sélectionnés sur milieu sélectif en présence de chloramphénicol, les plasmides ont été extraits puis contrôlés par PCR avec les

couples d'amorces PheR1f / PheR1r, PheR2f / PheR2r, Bad342Fw1 / Bad342Rv1 et par digestion enzymatique. Le plasmide pPR13 obtenu a été digéré par NdeI, traité à la Klenow Fragments, puis déphosphorylé par la SAP. Le plasmide pB33C3 est digéré par XmnI et NaeI. Le fragment d'ADN contenant la cassette [*araC*-P<sub>bad</sub> ::*colE3*] a été purifié sur gel. Les deux fragments d'ADN ont été ligués, puis transformés par électroporation dans *E. coli* 1661LIΔ*ara* électrocompétentes (selon protocole indiqué chapIII-1). Les transformants ont été sélectionnés sur milieu gélosé contenant du chloramphénicol, puis criblés en milieu liquide en présence d'arabinose. Les plasmides des transformants ayant le bon phénotype ont été extraits et contrôlés par digestion enzymatique. L'efficacité de la construction a été vérifiée par un test d'efficacité.

## 2.4 Modifications chromosomiques de *E.coli*

Les souches *E. coli* ont été modifiées par transduction de l'ADN d'une souche *E. coli* donatrice vers la souche *E. coli* réceptrice via le phage P1 (**Annexe 6**). Brièvement, 50 µL d'un stock de phage P1 ont été incubés sans agitation 15 min à 37°C en présence de 1 mL de culture bactérienne à D<sub>0</sub><sub>600nm</sub> 0,2-0,4 de la souche réceptrice dans un milieu LB contenant du chlorure de calcium (CaCl<sub>2</sub>). Quatre mL d'agar en surfusion ont été ajoutés. Le mélange a été déposé sur milieu nutritif, puis incubé 1 nuit à 37°C. La surcouche d'agar a été prélevée, les phages ont été extraits puis incubés 20 min à 37°C en présence de la souche réceptrice. Le mélange a été déposé sur milieu nutritif, puis incubé 1 nuit à 37°C. Les colonies obtenues ont été criblées en milieu solide suivant le phénotype attendu.

## 2.5 Test d'efficacité des constructions toxiques

Une colonie fraîche a été inoculée en milieu liquide contenant les antibiotiques adéquats. Après 12 h d'incubation à 37°C et 150 rpm, la culture bactérienne a été diluée en série au 1/10<sup>ème</sup> dans une solution de NaCl à 8 g.L<sup>-1</sup> jusqu'à la dilution 10<sup>-6</sup>. Cent µL de chacune des dilutions sont étalés sur boîte de Petri contenant du milieu gélosé contenant du chloramphénicol avec ou sans arabinose. Après incubation 1 nuit à 37°C, les colonies (UFC) ont été dénombrées. Un taux d'échappement a été calculé en divisant le nombre d'UFC.mL<sup>-1</sup> se développant sur milieu avec inducteur par le nombre d'UFC.mL<sup>-1</sup> se développant sur milieu sans inducteur.

## 2.6 Extraction de l'ADN métagénomique

Un échantillon de biofilm a été prélevé à partir d'un bioréacteur anaérobie de traitement d'eaux usées d'une cokerie (Shanghai, China) suivant le protocole indiqué dans Liu *et al.* (2006) (Liu *et al.*, 2006). L'ADN a été extrait à l'aide du kit FastDNA SPIN for soil (Bio101 systems, MP biomedical, Illkirch, France) puis stocké à -20°C.

## 2.7 Transformation / recombinaison ADN dans la souche réceptrice

La préparation des cellules pour la transformation et la recombinaison a été effectuée selon le protocole décrit par Thomason *et al.* (2005) (Thomason *et al.*, 2005). Brièvement, une préculture d'une nuit de la souche réceptrice *E. coli* 1661LIΔara (pSIM6, pPRE), suivie d'une culture jusqu'à  $D_{0.600nm}$  0,4-0,6 ont été effectuées à 30°C en présence des pressions de sélection ampicilline et chloramphénicol. La culture a été placée 15 min à 42°C avec agitation, puis 5 min à 4°C avec agitation douce. La souche réceptrice a été rendue électrocompétente par une succession de lavages dans du glycérol 10 % froid selon le protocole indiqué dans le sous chapitre III-1. Cent  $\mu$ L de la souche réceptrice rendue recombinogène et électrocompétente ont été mis en présence de la solution d'ADN à ajustée à 10  $\mu$ L avec de l'eau stérile dans une cuve d'électroporation de 2 mm (BTX, MP Biomedicals, Illkirch, France) pré-refroidie à 4°C. Un choc électrique de 2,5 kV.cm<sup>-1</sup>, 200  $\Omega$ , 25  $\mu$ F a été appliqué. Du milieu LB a été ajouté pour QSP 1 mL, puis les cellules ont été régénérées par incubation 2 h à 30°C avec agitation douce. Les transformants ont été dilués au 1/10<sup>ème</sup> en série dans du NaCl à 8 g.L<sup>-1</sup>, étalés sur milieu solide sélectif contenant ou non de l'arabinose, puis incubés une nuit à 30°C.

Dans une variante du protocole, 900  $\mu$ L de la suspension bactérienne régénérée ont été dilués dans 10 mL de milieu LB contenant du chloramphénicol et de l'ampicilline 5 h à 30°C. La culture a été diluée au 1/100<sup>ème</sup> dans du milieu LB contenant du chloramphénicol puis incubée 1 nuit à 37°C. Les plasmides ont été extraits puis retransformés dans *E. coli* 1661LIΔara électrocompétentes. Les transformants ont été dilués au 1/10<sup>ème</sup> en série dans du NaCl à 8 g.L<sup>-1</sup>, étalés sur milieu solide sélectif contenant ou non de l'arabinose, puis incubés 1 nuit à 37°C.

## 2.8 Analyses de séquences

Les similarités entre séquences ADN ont été vérifiées par analyse BLAST (*Basic Local Alignment Search Tool*) (Altschul *et al.*, 1990) via le serveur NCBI (*National Center for Biotechnological Information*) (<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi>). Les sites de recombinaison ont été définis par alignement de séquences multiples avec le logiciel clustalX (Chenna *et al.*, 2003).

## 3 RESULTATS ET DISCUSSION

### 3.1 Définition et modifications de la souche réceptrice :

L'intégration génomique de fragments d'ADN linéaires par recombinaison dans les souches *recBCD* est stimulée par la présence de sites *chi*, qui diminuent l'activité exonucléasique de RecD et favorisent la fixation de RecA sur l'ADN simple brin (ADNsb). Ces sites *chi* ou des sites similaires (*chi*-like) sont présents chez de nombreux organismes (El Karoui *et al.*, 1999). La présence de deux sites *chi* à chacune des extrémités du fragment linéaire et la restriction enzymatique à proximité de ces sites accroissent l'activité recombinogène (Dabert & Smith, 1997). Leur absence conduit à la diminution de la probabilité d'intégration dans les souches *recBCD*.

Une souche *E. coli* mutée *recD*, *E. coli* 1661, a donc été choisie pour son utilisation comme souche réceptrice. La mutation du gène *recD* entraîne une perte de l'activité exonucléasique tout en conservant l'activité recombinogène et la viabilité des cellules (Amundsen *et al.*, 1986). Les mutants *recD* deviennent (hyper) recombinogènes en l'absence de sites *chi* (Chaudhury & Smith, 1984; Dabert & Smith, 1997). La cassette [P3::*immE3-kn<sup>r</sup>-lacI<sup>f</sup>*] contenue dans *E. coli* TELI a été transduite dans *E. coli* 1661 via le phage P1, et les transductants sélectionnés sur milieu contenant de la kanamycine. La souche *E. coli* 1661LI obtenue ne se développe pas sur milieu contenant du tryptophane, indiquant l'insertion de la cassette dans la région *trp* du chromosome de *E. coli* 1661 comme attendu. Contrairement aux souches *E. coli immE3*-, *E. coli* 1661LI se développe sur milieu contenant de la colicine. De plus, le plasmide pB33C3 contenant le gène *colE3* a pu être transformé avec succès dans *E. coli* 1661LI. Le gène codant l'antitoxine ImmE3 est donc exprimé, confirmant la fonctionnalité de P3::*immE3* dans *E. coli* 1661LI.

**Tableau 9 : Efficacité de pB33C3 en fonction de la quantité d'arabinose disponible pour l'induction de l'expression de *colE3***

Plasmide	Souche <i>E.coli</i>	C (arabinose) (g.L <sup>-1</sup> )	Nombre de reproduction	Efficacité moyenne	Ecart type
pB33C3	<b>1661LI</b>	10	2	8,40E-06	4,77E-06
pB33C3	<b>1661LI</b>	40	1	9,06E-07	
pB33C3	<b>1661LIΔara</b>	10	3	1,83E-06	9,76E-07

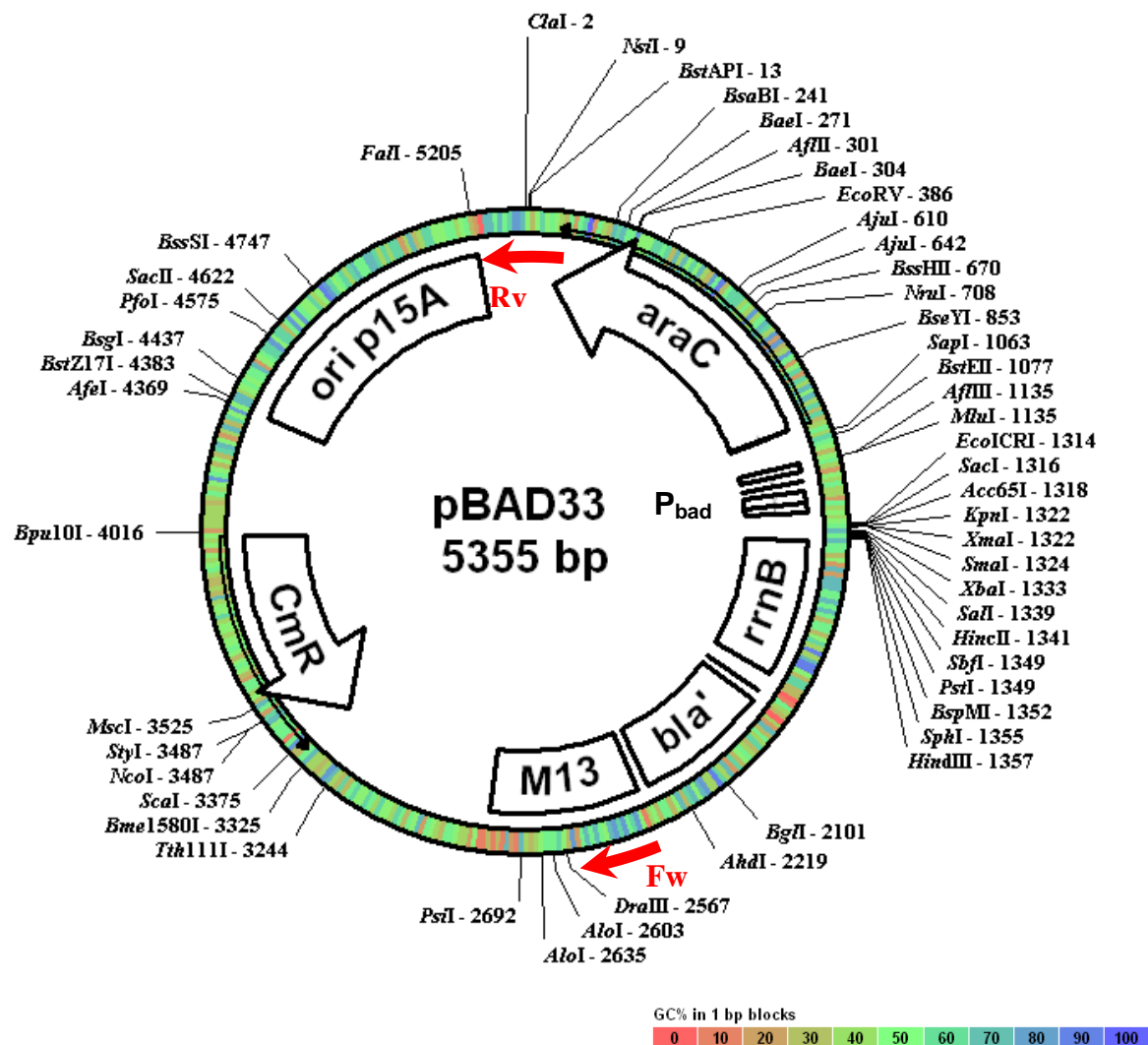
L'efficacité du plasmide pB33C3 précédemment construit (sous chapitre III-1) a été testée dans la souche *E. coli* 1661LI obtenue. Le test d'induction en milieu solide indique un taux d'échappement pour *E. coli* 1661LI (pB33C3) de  $10^{-5}$  identique à celui obtenu pour *E. coli* TELI (pB33C3). En augmentant la concentration en arabinose à 4 % nous avons constaté un taux d'échappement 10 fois plus faible. Des expériences menées par Siegele *et al.* (1997) ont montré que seule une partie de la population est induite si les concentrations en inducteur utilisées sont inférieures aux concentrations saturantes (Siegele & Hu, 1997). La quantité d'arabinose utilisée ici est cependant bien supérieure à cette concentration seuil.

Une explication possible de cette diminution du taux d'échappement avec l'augmentation de la concentration en arabinose est la métabolisation de l'inducteur. *E. coli* TELI et *E. coli* 1661LI sont en effet toutes deux capables de métaboliser l'arabinose (**Tableau 9**). L'arabinose alors métabolisé ne serait plus disponible pour l'induction, ce qui n'entraînerait la mort que d'une partie de la population.

Pour confirmer cette hypothèse, la souche *E. coli* 1661LI a été mutée dans son opéron arabinose. La cassette  $\Delta araBAD$  de la souche *E. coli* BW25141 a été transduite par le phage P1 dans une souche de *E. coli* 1661LI, préalablement rendue auxotrophe pour la leucine (leu-) et tétracycline résistante. L'insertion de la cassette a été vérifiée par le rétablissement du phénotype leu+ (croissance en présence de leucine), la perte de la résistance à la tétracycline et l'absence de croissance de la souche en présence d'arabinose. L'efficacité de pB33C3 a été testée dans la souche nouvellement construite pour une concentration finale en arabinose de 1 %. Le taux d'échappement obtenu est de  $10^{-6}$ , soit 10 fois plus faible que la souche non mutée. L'absence de métabolisation de l'arabinose, liée à l'inactivation de l'opéron arabinose, améliore d'un ordre de grandeur l'efficacité tueuse liée à l'expression du gène *colE3*.

### **3.2 Définition du vecteur pour le système de contre sélection positive**

Un nouveau système de contrôle de l'expression inductible du gène *colE3* a précédemment été développé (sous chapitre III-1), constitué du système de régulation dérivé de l'opéron arabinose, complété de l'expression constitutive à faible taux du gène *immE3*. Le système développé est fonctionnel sur le plasmide pBAD33, portant une origine de réplication p15A. Des expérimentations futures seront réalisées pour évaluer ce système de régulation dans des vecteurs à plus faible nombre de copies, comme les plasmides dérivés de pSC101. Dans cette étude, un vecteur dérivé de pBAD33 a été utilisé pour le développement du système de contre sélection positive.



**Figure 9 : Emplacement des amorces Bad33bglIIIFw2 et Bad33pstIRv2 sur pBAD33 pour l'amplification de pB342**

Dessin réalisé avec le logiciel pDRAW d'après la séquence nucléotidique reconstruite de pBAD33 selon les indications de constructions génétiques données par les auteurs (Guzman *et al.*, 1995). *araC* : gène répresseur, *Bla'* : gène *bla* partiellement délété, *CmR* : gène de résistance au chloramphénicol, *Ori p15A* : origine de réplication p15A, *Pbad* : promoteur, *rrnB* : terminateur. En rouge : les amorces avec Fw correspondant à l'amorce sens Bad33bglIIIFw2, Rv correspondant à l'amorce antisens Bad33pstIRv2.



L'analyse par BLAST de pBAD33 révèle cependant des zones de similarités du plasmide avec la cassette [P3 ::*immE3-kn'*-*lacI<sup>q</sup>*] transférée dans le chromosome de la souche hôte *E. coli*. Trois-cent trente huit nucléotides en amont et 208 nucléotides en aval de *immE3* sont similaires à une région de pBAD33 comprise entre le terminateur *rrnB* et l'Origine M13 (**Figure 9**). Cette région correspond à l'ancien emplacement du gène *bla*, gène conférant la résistance à l'ampicilline, qui a été tronqué au cours des étapes de construction de pBAD33 (Guzman *et al.*, 1995). Bien que le plasmide ne confère plus le phénotype de résistance à l'ampicilline, plus de 400 nucléotides de l'ORF *bla* sont encore présents sur le plasmide.

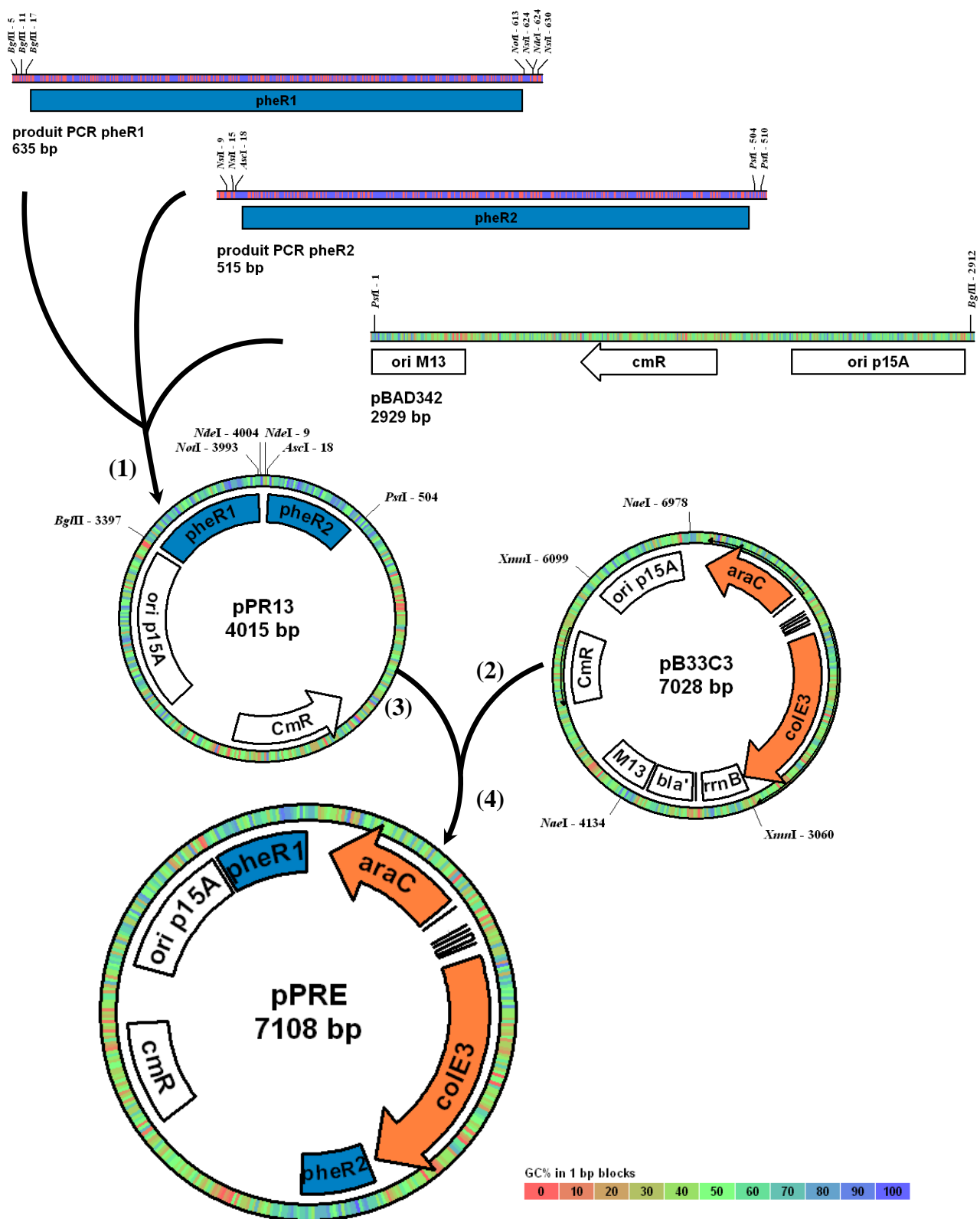
La présence de ces régions d'ADN résiduelles peut être problématique dans le cadre du développement d'un outil nécessitant l'optimisation des potentialités recombino-gène de l'espèce bactérienne utilisée. La recombinaison pouvant être amorcée pour une vingtaine de paires de bases (Shen & Huang, 1986), toute zone de forte similarité de séquences entre le plasmide et le chromosome pourrait entraîner des événements de recombinaison non contrôlés, risquant d'altérer l'efficacité de contre sélection. Cette séquence  $\Delta bla$  a donc été éliminée du vecteur portant le système de contre sélection.

Pour ce faire, le couple d'amorces Bad33BglIIFw2 / Bad33PstIRv2 a été défini afin d'amplifier la séquence nucléotidique de pBAD33 en aval de la zone d'homologie et en amont du gène *araC* (**Figure 9**). Le nouveau vecteur pB342 amplifié ne contient ni séquence  $\Delta bla$ , ni le gène *araC*. Des sites de restriction BglII et PstI ont été ajoutés dans les amorces pour permettre le clonage des sites de recombinaison pheR1 et pheR2 dans pB342.

### 3.3 Définition des zones de recombinaison

Les zones de recombinaison ont été définies en fonction des régions conservées des gènes ciblés. Nous nous sommes intéressés aux gènes codant les phénols hydroxylases, enzymes intervenant dans le catabolisme de composés aromatiques phénoliques (Hughes & Bayly, 1983; Zhang *et al.*, 2004). L'alignement de séquences nucléotidiques sous clustalX de gènes provenant de 3 espèces bactériennes (**Annexe 8**) a permis l'identification de deux régions conservées.

Deux couples d'amorces ont été définies pour l'amplification de chacune de ces deux régions conservées : PheR1f / PheR1r pour l'amplification de la première zone de recombinaison pheR1 d'environ 600 pb (589 pb chez *Ralstonia metallidurans* CH34) et PheR2f / PheR2r pour l'amplification de la seconde zone de recombinaison pheR2 d'environ 500 pb (475 pb chez *R. metallidurans* CH34). Des sites de restriction ont été ajoutés aux amorces pour orienter le clonage des deux produits d'amplification dans le plasmide pB342



**Figure 10 : Stratégie de construction du plasmide pPRE**

(1) génération de bouts francs pour les produits PCR pheR1 et pheR2, digestion des trois produits PCR par PstI, BglII, puis ligation pour obtenir pPR13 (2) digestion de pB33C3 par XmnI et NaeI, purification de la cassette contenant *colE3* (3) digestion de pPR13 par NdeI et génération de bouts francs (4) ligation de la cassette contenant *colE3* dans pPR13 pour la construction de pPRE. En bleu : les sites de recombinaison, en orange : la cassette toxique. (Dessins non à l'échelle)

(**Figure 10**). Ainsi, les sites de restriction reconnus par les enzymes BglIII et PstI ont été ajoutés aux amorces PheR1f et PheR2r respectivement, et le site de restriction AvaIII a été ajouté aux amorces PheR1r et PheR2f. Le choix des enzymes de restriction a été contraint par l'absence de ces mêmes sites de restriction dans le vecteur et la cassette toxique utilisés.

### 3.4 Construction du système de contre sélection positive

L'ADN métagénomique, extrait de biofilms bactériens d'un bioréacteur anaérobie de traitements d'eaux usées d'une cokerie, a servi de cible pour l'amplification des sites de recombinaison pheR1 et pheR2 (les conditions d'amplification utilisées sont reportées en **Annexe 9**). Les produits PCR ont été digérés et clonés dans pB342. Après transformation des produits de ligation dans *E. coli* DH5 $\alpha$ , les plasmides extraits des transformants obtenus contiennent uniquement le site de recombinaison pheR1, à cause de la complémentarité des séquences entre les extrémités cohésives générées par AvaIII et PstI. Les sites de restriction AvaIII ont donc été remplacés par des extrémités en bouts francs pour le clonage orienté de pheR1 et pheR2 dans pB342. Après transformation des produits de ligation dans *E. coli* DH5 $\alpha$ , le plasmide pPR13 extrait d'un des transformants obtenus possède le profil de restriction attendu. Les sites de recombinaison clonés ont été amplifiés par le couple d'amorces Bad342Fw1 / Bad342Rv1 puis séquencés. Les séquences nucléotidiques de pheR1 et pheR2 sont données dans l'**Annexe 10**.

Pour finaliser la construction du système de contre sélection positive, la cassette toxique contenant *colE3* a été clonée entre les sites de recombinaison du plasmide pPR13. Pour ce faire, le fragment d'ADN de pB33C3 digéré par XmnI et NaeI contenant la cassette [*araC*-P<sub>bad</sub>::*colE3*] a été cloné au niveau du site de restriction NdeI de pPR13 par clonage en bout franc, puis transformé dans *E. coli* 1661LI $\Delta$ ara. Les transformants obtenus ont été criblés pour leur sensibilité à l'arabinose. Un clone *E. coli* possédant le plasmide pPRE attendu (**Figure 10**) a été isolé. L'efficacité de la construction a été vérifiée par induction en milieu solide. Le taux d'échappement est de 10<sup>-6</sup>.

Le plasmide pSIM6, contenant le système red gam sous le contrôle du promoteur P<sub>L</sub> inductible à la température, a été introduit par transformation dans la souche réceptrice *E. coli* 1661LI $\Delta$ ara (pPRE). Ce dernier possède une origine de répllication compatible avec pPRE, permettant ainsi la présence simultanée des deux plasmides dans la souche réceptrice. La souche réceptrice peut être guérie de pSIM6 par incubation 8 h à 37°C en absence de pression de sélection (Datta *et al.*, 2006).

### 3.5 Efficacité du système de contre sélection à piéger l'ADN ciblé

Dans le but de vérifier qu'un système de contre sélection positive peut capturer de l'ADN métagénomique, des premiers tests expérimentaux ont été conduits *in vitro* avec une souche réceptrice *E. coli*. Pour ces premiers tests expérimentaux, une cassette toxique simple constituée du seul gène toxique *colE3*, encadrée des sites de recombinaison *pheR1* et *pheR2*, a été utilisée. Suivant la définition des sites de recombinaison, un fragment d'ADN métagénomique d'environ 2 kb (cas de *R. metallidurans*) devrait remplacer la cassette [*pheR1-araC-P<sub>bad</sub>::colE3-pheR2*] (4,6 kb en taille) par double *crossing over*.

L'ADN métagénomique extrait de biofilms a été utilisé pour électrotransformer la souche *E. coli* 1661LIΔara (pPRE, pSIM6) rendue recombinogène et électrocompétente. Les cellules transformées et soumises à la recombinaison ont été étalées sur milieu sélectif contenant de l'arabinose pour induire l'expression de *colE3*. Ainsi les souches réceptrices, n'ayant pas intégré l'ADN métagénomique par double *crossing-over* au niveau des sites *pheR1* / *pheR2*, doivent être tuées par les produits exprimés par le gène *colE3*. Les clones recombinants peuvent ainsi être spécifiquement détectés dans la limite d'efficacité du système de contre sélection, définie par le taux d'échappement à  $10^{-6}$  de la cassette toxique.

Sur près de  $10^8$  UFC.mL<sup>-1</sup> de souches réceptrices soumises à un choc électrique en présence d'ADN métagénomique, quelques colonies sont capables de se développer sur milieux contenant de l'arabinose, et ce avec une fréquence de l'ordre de  $10^{-5}$ . La fréquence d'obtention de ces colonies est ainsi 10 fois supérieure à la fréquence d'apparition de clones résistants à l'expression de la cassette toxique (fréquence obtenue pour les conditions utilisées dans le test d'efficacité). Néanmoins, un témoin négatif réalisé avec de l'eau remplaçant l'ADN métagénomique a donné les mêmes résultats. Ces colonies capables de se développer en présence d'arabinose ne seraient donc pas des souches réceptrices recombinées mais des faux positifs, c'est-à-dire des souches ayant échappé à l'expression de la cassette toxique. Le taux d'échappement dans ces nouvelles conditions expérimentales serait ainsi 10 fois plus important, déplaçant ainsi le seuil de détection des souches recombinées par l'outil de capture présenté d'un ordre de grandeur.

Afin de comprendre l'origine de l'augmentation du taux d'échappement, les plasmides des clones obtenus ont été extraits puis analysés par PCR avec le couple d'amorces *PheR1f* / *PheR2r* (à une température d'hybridation *T<sub>m</sub>* de 52°C). Ce couple amplifie la région d'ADN contenant les sites de recombinaison *pheR1* et *pheR2*. Les plasmides analysés (au nombre de 8) présentent tous un fragment d'ADN dont la taille est inférieure à celle correspondant à la

seule présence de [pheR1-pheR2] (1,1 kb). Des événements mutationnels de délétion de la cassette toxique expliqueraient donc l'augmentation du taux d'échappement observé, cette délétion étant favorisée par l'activation de la recombinaison

### 3.6 Vérification d'un piégeage effectif de l'ADN par recombinaison

L'emploi de pSIM6 devrait permettre la détection de recombinants (vrais positifs) à des fréquences pouvant atteindre de  $2,9 \cdot 10^{-4}$  dans *E. coli* (Datta *et al.*, 2006). Dans nos conditions expérimentales, la fréquence d'apparition de ces recombinants apparaît cependant inférieure à  $10^{-5}$ , puisque aucun clone recombinant n'a pu être isolé.

Ce phénomène peut être expliqué par la dilution de l'ADN recherché dans l'ADN métagénomique (Bertolla *et al.*, 2000; de Vries *et al.*, 2001; Kay *et al.*, 2002). Pour confirmer cette hypothèse, l'ADN métagénomique ciblé a été concentré par une étape PCR : le couple d'amorces PheR1f / PheR2r a été utilisé pour l'amplification de l'ADN métagénomique présentant une similarité de séquences avec les deux amorces. Le produit PCR de 2 kb obtenu a servi aux essais de transformation / recombinaison dans la souche réceptrice *E. coli* 1661LIΔara (pPRE, pSIM6). Aucun clone recombinant n'a cependant pu être détecté. Ceci peut être expliqué par le phénotype de la souche réceptrice utilisée. Cette dernière en effet n'est pas mutée dans le gène *hsdR*, dont les produits d'expression interviennent dans la restriction de l'ADN. Cette souche dégraderait donc l'ADN ne portant pas la méthylation adéquate, tels les produits PCR (Tock & Dryden, 2005).

Dans une seconde hypothèse, l'absence de détection de clones recombinants serait liée aux multiples copies du système de contre sélection. Le plasmide pPRE, possédant une origine de répllication p15A, est présent en 15-20 copies par cellule (Chang & Cohen, 1978). Même si un événement de remplacement de la cassette toxique par le fragment ADN ciblé a lieu sur une des copies du plasmide, les autres copies portent toujours la cassette toxique. La question se pose alors de la ségrégation des différentes copies du plasmide au cours de la division cellulaire (Nordstrom & Austin, 1989), et donc l'uniformisation de l'information initialement présente sur une des copies à l'ensemble des copies plasmidiques d'une cellule au cours des générations cellulaires. Les études portant sur la recombinaison sont généralement effectuées au niveau chromosomique (Datsenko & Wanner, 2000; de Vries *et al.*, 2001; Murphy, 1998). Certaines études ont également été réalisées sur plasmide, cependant l'ADN intégré apportait une sélection positive telle que la résistance à un antibiotique (Kay *et al.*, 2002; Nielsen *et al.*, 2000; Yu *et al.*, 2000). Dans ces conditions, l'intégration de l'ADN sur une copie de plasmide suffit à conférer la résistance à l'antibiotique (pour une concentration

donnée). Dans notre cas, le maintien d'une seule copie non recombinée suffirait à entraîner la mort de la cellule hôte.

Pour vérifier cette deuxième hypothèse, les plasmides des transformants ont été extraits puis retransformés dans *E. coli* 1661LIΔara. De telle manière, les plasmides présents dans une cellule sont redistribués, évitant ainsi la présence conjointe dans une souche réceptrice d'un plasmide ayant capturé l'ADN ciblé avec un plasmide de contre sélection intact. Cependant aucune souche réceptrice possédant un plasmide ayant capturé l'ADN n'a pu être détecté. La détection de souches réceptrices recombinées a nécessité non seulement l'extraction des plasmides, mais également la digestion du plasmide par l'enzyme SacI coupant à l'intérieur de la cassette toxique avant leur réintroduction par transformation dans la souche *E. coli* DH5α non immune (immE3-).

Six transformants ont ainsi été obtenus lorsque de l'ADN métagénomique a été utilisé pour la transformation / recombinaison, et seulement 3 transformants lorsque l'ADN métagénomique a été remplacé par de l'eau. L'analyse de ces transformants par amplification PCR indique, pour les trois transformants du témoin eau, la délétion d'une partie des sites de recombinaison, témoignant des événements de recombinaison entre les deux sites de recombinaison. Pour les essais avec l'ADN métagénomique, 3 transformants montrent la présence d'un insert de taille supérieure à celle correspondant à la cassette [pheR1-*araC*-P<sub>bad</sub>::*colE3*-PheR2]. Il semblerait donc que sur 10<sup>8</sup> UFC.mL<sup>-1</sup> de souches réceptrices soumises à la transformation / recombinaison, au moins un événement de recombinaison / intégration d'un fragment d'ADN eut lieu au niveau des sites pheR1 / pheR2. La fréquence de cet événement serait environ de 10<sup>-7</sup>, soit inférieure au seuil de détection du système de contre sélection développé. La taille de l'ADN inséré est cependant bien supérieure à celle attendue, le séquençage du fragment intégré confirmera le piégeage effectif d'un ADN ciblé.

## **4 CONCLUSION**

Un système de contre sélection positive a été développé, et des premiers tests expérimentaux ont été réalisés afin d'évaluer la possibilité d'un piégeage de l'ADN de la microflore tellurique par l'outil de capture. Ces premiers tests expérimentaux ont notamment mis en valeur les limites de l'outil développé et les conditions requises pour un piégeage effectif de l'ADN de la microflore tellurique.

Les événements de transformation / recombinaison d'ADN ciblé provenant du métagénome semblent avoir lieu à des fréquences de  $10^{-7}$ , aussi une cassette toxique dont le taux d'échappement est inférieur à  $10^{-7}$  est nécessaire pour la détection de souches recombinées. Par ailleurs, la présence de similarités entre les deux sites de recombinaison peut altérer l'efficacité du système de contre sélection à détecter les seules souches réceptrices recombinées. Dans les travaux présentés dans ce sous chapitre, ces événements de recombinaison non attendus ont provoqué une perte d'un ordre de grandeur dans l'efficacité du système de contre sélection. Un soin particulier doit donc être apporté pour la définition des sites de recombinaison, en évitant toute similarité entre les deux sites de recombinaison.

Ces travaux ont par ailleurs soulevé d'intéressantes questions quant à la ségrégation des plasmides dans les cellules filles et la possibilité d'une uniformisation de l'information génétique contenue sur une des copies à l'ensemble des copies plasmidiques au cours des générations cellulaires. De telles données pourraient influencer le choix du vecteur portant le système de contre sélection. Le système, actuellement développé sur un vecteur à moyen nombre de copies, pourrait être transféré sur un vecteur à une seule copie par cellule si une uniformisation de l'information génétique portée par les vecteurs dérivés p15A n'a pu être démontrée.

Les expérimentations effectuées dans ce sous chapitre n'ont pu confirmer ou infirmer cette uniformisation, puisque la seule extraction / réintroduction des plasmides dans une souche réceptrice n'a pu suffire à la détection de souches recombinées. La restriction enzymatique et l'emploi d'une souche non immune ont été nécessaires pour éliminer la majeure partie des transformants faux positifs, c'est-à-dire des souches réceptrices n'ayant pas intégré d'ADN par échange allélique au niveau des sites de recombinaison.

Des travaux complémentaires sont donc nécessaires pour vérifier cette uniformisation, ainsi que pour développer une cassette toxique encore plus efficace.

## **SOUS-CHAPITRE III-3 :**

### **OPTIMISATION DE L'EFFICACITE DE LA CASSETTE TOXIQUE DU SYSTEME DE CONTRE SELECTION POSITIVE POUR LA DETECTION D'EVENEMENTS RARES**

#### **1 INTRODUCTION**

Les premiers travaux d'application de l'outil de capture de gènes présents dans le métagénome (sous chapitre III-2) sont révélateurs des potentialités et limites de la technologie en développement. Rappelons que dans les cas les plus favorables de transformation *in vitro* d'une souche *E. coli* équipée du système lambda red et de deux régions de 1 000 pb de totale similarité avec l'ADN donneur pour amorcer la recombinaison, la fréquence de recombinaison d'un ADN linéaire peut atteindre des valeurs très élevées de l'ordre de  $3.10^{-4}$  (Murphy, 1998). Pour les applications proposées ici, il apparaît très clairement que les fréquences de transformation et recombinaison demeureront beaucoup plus faibles du fait de la dilution des gènes ciblés au sein du métagénome (Bertolla *et al.*, 2000; de Vries *et al.*, 2001; Kay *et al.*, 2002). Ces fréquences seront également d'autant plus faibles que la similarité de séquences entre l'ADN pénétrant et les régions cibles présentes dans la souche réceptrice diminue (Vulic *et al.*, 1999).

Les résultats présentés précédemment indiquent que la transformation de la souche réceptrice avec l'ADN du métagénome ne permettra la production de clones recombinants qu'à des fréquences inférieures à  $10^{-6}$ . Ces données signifient que de tels clones ne pourront être détectés que si le système de contre sélection présente un seuil de détection inférieur à cette valeur. Or les travaux préliminaires basés sur la seule activité létale de *colE3* dont le taux d'échappement avoisine  $10^{-6}$  montrent bien la nécessité d'un système basé sur la combinaison de deux gènes toxiques de façon à baisser le taux d'échappement.

Des systèmes développés pour l'endiguement génétique de plasmides montrent un taux d'échappement atteignant  $10^{-8}$  par la duplication d'un même gène toxique (Knudsen *et al.*, 1995; Knudsen & Karlstrom, 1991). Ces constructions souffrent cependant de limitations liés à des événements de recombinaison homologue entre les deux copies identiques du gène toxique (Knudsen *et al.*, 1995, sous chapitre III-4) ou à l'inactivation du système régulateur



**Tableau 10 : Liste et caractéristiques des plasmides utilisés ou construits**

Plasmide	Commentaires	Génotype	Phénotype	Référence
<b>pBAD33</b>	Vecteur d'expression	<i>cm<sup>r</sup></i> , ori p15A, <i>araC</i> , <i>P<sub>bad</sub></i>	Cm <sup>r</sup>	Guzman <i>et al.</i> , 1995
<b>pB341</b>	Vecteur de clonage	<i>cm<sup>r</sup></i> , ori p15A	Cm <sup>r</sup>	Cette étude
<b>pSK360</b>	Cible pour l'amplification de la cassette [ <i>P<sub>A1-O3/O4::relF-tc<sup>r</sup></sub></i> ]	<i>amp<sup>r</sup></i> , <i>tc<sup>r</sup></i> , 2 x ( <i>P<sub>A1-O3/O4::relF</sub></i> ), ori pMB1	Amp <sup>r</sup> , Tc <sup>r</sup> , IPTG <sup>s</sup>	Knudsen <i>et al.</i> , 1995
<b>pB33C3</b>	Cible pour l'amplification de la cassette [ <i>araC-P<sub>bad::colE3-rrnB</sub></i> ]	<i>cm<sup>r</sup></i> , ori p15A, <i>araC-P<sub>bad::colE3</sub></i>	Cm <sup>r</sup> Ara <sup>s</sup>	Sous chapitre III-1
<b>pB33C3RTc3G6</b>	Clonage de la cassette [ <i>P<sub>A1-O3/O4::relF-tc<sup>r</sup></sub></i> ] dans pB33C3	<i>cm<sup>r</sup></i> , ori p15A, <i>araC-[P<sub>A1-O3/O4::relF-tc<sup>r</sup>]-P<sub>bad::colE3</sub></sub></i>	Cm <sup>r</sup> Tc <sup>r</sup> , Ara <sup>r</sup> , IPTG <sup>s</sup>	Cette étude
<b>pB341C3</b>	Clonage de la cassette [ <i>araC-P<sub>bad::colE3-rrnB</sub></i> ] dans pB341	<i>cm<sup>r</sup></i> , ori p15A, <i>araC-P<sub>bad::colE3</sub></i>	Cm <sup>r</sup> Ara <sup>s</sup>	Cette étude
<b>pB341RTc</b>	Clonage de la cassette [ <i>P<sub>A1-O3/O4::relF-tc<sup>r</sup></sub></i> ] dans pB341	<i>cm<sup>r</sup></i> , ori p15A, <i>P<sub>A1-O3/O4::relF-tc<sup>r</sup></sub></i>	Cm <sup>r</sup> Tc <sup>r</sup> IPTG <sup>s</sup>	Cette étude
<b>pB341C3RTc1</b>	Clonage de la cassette [ <i>P<sub>A1-O3/O4::relF-tc<sup>r</sup></sub></i> ] dans pB341C3	<i>cm<sup>r</sup></i> , ori p15A, <i>araC-P<sub>bad::colE3</sub> [P<sub>A1-O3/O4::relF-tc<sup>r</sup>]</sub></i>	Cm <sup>r</sup> Tc <sup>r</sup> IPTG <sup>r</sup>	Cette étude
<b>pB341C3RTc2</b>	Clonage de la cassette [ <i>P<sub>A1-O3/O4::relF-tc<sup>r</sup></sub></i> ] dans pB341C3	<i>cm<sup>r</sup></i> , ori p15A, <i>araC-P<sub>bad::colE3</sub>, P<sub>A1-O3/O4::relF-tc<sup>r</sup></sub></i>	Cm <sup>r</sup> Tc <sup>r</sup> IPTG <sup>s</sup>	Cette étude

Amp<sup>r</sup> : résistant à l'ampicilline, Ara<sup>r</sup> : résistant à l'induction par l'arabinose, Ara<sup>s</sup> : sensible à l'induction par l'arabinose, Cm<sup>r</sup> : résistant au chloramphénicol, IPTG<sup>r</sup> : résistant à l'induction par l'IPTG, IPTG<sup>s</sup> : sensible à l'induction par l'IPTG, Tc<sup>r</sup> : résistant à la tétracycline.

contrôlant l'expression des deux gènes toxiques (Knudsen *et al.*, 1995). C'est pour limiter ces biais que l'utilisation de deux gènes dont les produits visent des cibles cellulaires différentes, chacun contrôlé par un système régulateur autonome, est préconisée (Torres *et al.*, 2003).

Un système toxique possédant deux gènes toxiques différents a déjà été rapporté (Torres *et al.*, 2003). Ce système a cependant été construit dans le cadre d'un endiguement génétique. Ainsi les gènes d'immunité *ecoRIM* et *immE3* ont été placés sur le chromosome de la souche hôte, sous le contrôle de promoteurs constitutifs forts. Les deux gènes toxiques correspondants *ecoRIR* et *colE3* ont été clonés sur un plasmide, l'un d'entre eux ayant été cloné sous le contrôle d'un promoteur constitutif permettant l'expression continue de la protéine toxique (Torres *et al.*, 2003). Un tel système est intéressant pour éviter la dissémination de l'information génétique portée par le plasmide, mais n'est pas exploitable tel quel dans notre étude où les gènes toxiques doivent être placés dans un système d'expression inductible.

Pour construire une cassette toxique inductible, deux gènes aux fonctions toxiques ciblant des composants cellulaires différents ont été clonés sur un même plasmide, l'expression de chacun des gènes étant régulée indépendamment. L'objectif de ces travaux était la construction d'une cassette toxique efficace qui puisse être facilement mobilisée pour des usages ultérieurs, notamment pour son insertion entre les régions cibles dévolues à la recombinaison. Le choix des gènes toxiques, les stratégies de clonage des deux gènes sur un même plasmide et les résultats d'efficacité tueuse des constructions toxiques obtenues sont présentés dans ce sous chapitre.

## **2 MATERIELS ET METHODES**

### **2.1 Bactéries et plasmides**

Les souches de *E. coli* 1661LIΔara (*F*- *lac*-3350 *galK2 galT22 recD IN(rrnD-rrnE)1 rpsL179 Δtrp::[kn<sup>r</sup>-immE3-lacI<sup>q</sup>] ΔaraBAD*) (Sous chapitre III-2), *E. coli* DH5α (*F*- *φ80lacZΔM15 Δ(lacZYA-argF)U169 deoR recA1 endA1 hsdR15(rk-, mk+) phoA supE44 thi-1 gyrA96 relA1 l*) (InvitroGen, Cergy Pontoise) et *E. coli* DH5αZ1 (*F*- *φ80lacZΔM15 Δ(lacZYA-argF)U169 deoR recA1 endA1 hsdR15(rk-, mk+) phoA supE44 thi-1 gyrA96 relA1 l- spr lacIq tetR*) (Lutz & Bujard, 1997) ont été conservées dans du glycérol (Euromedex, Souffelweyersheim, France) 25 % en concentration finale et stockées à -80°C.

**Tableau 11 : Liste des amorces, de leurs caractéristiques et des conditions d'amplification utilisées**

Amorce	Séquence en 5'-	ADN cible	Site restriction	Gène amplifié	Mésappariements <sup>b</sup>	Tm <sup>a</sup> (°C)
<b>Bad33ascIFw3</b>	TTAGATTGATTGGCGCGCCCTGTAGC	pBAD33	SgsI	Ori p15A, <i>cm<sup>r</sup></i>	2	66
<b>Bad33notIRv2</b>	GCAGATCCTCTACGGCGGCCGCATC	pBAD33	NotI	Ori p15A, <i>cm<sup>r</sup></i>	3	66
<b>PareIFTcascIFw2</b>	GTTGAGGAGGGCGCGCCACATCAGAG	pSK360	SgsI	P <sub>A1-O3/O4</sub> :: <i>relF-tc<sup>r</sup></i>	3	63
<b>PareIFTcndeIRv5</b>	CAATTATCTCTTCATATGTAGCACCTG	pSK360	NdeI	P <sub>A1-O3/O4</sub> :: <i>relF-tc<sup>r</sup></i>	1	63
<b>PbcolE3NotIFw5</b>	GATCTGCTCATGTGCGGCCGCTTATCA	pB33C3	NotI	<i>araC</i> -P <sub>bad</sub> :: <i>colE3</i> -rrnB	4	59
<b>PbcolE3ndeIRv5</b>	CAAAAGCATATGTAGAAACGCAAAAAGGC	pB33C3	NdeI	<i>araC</i> -P <sub>bad</sub> :: <i>colE3</i> -rrnB	3	59
<b>PareFTcSgsIFw1</b>	GCTCCTGAAAGGCGCGCCAAGCTAG	pSK360	SgsI	P <sub>A1-O3/O4</sub> :: <i>relF-tc<sup>r</sup></i>	3	59
<b>PareFTcSgsIRv2</b>	CCTGAAGGGCGCGCCATACGATATAAGTTG	pSK360	SgsI	P <sub>A1-O3/O4</sub> :: <i>relF-tc<sup>r</sup></i>	4	59

Les sites de restrictions sont indiqués sur la séquence par le soulignement

<sup>a</sup>Tm : température d'hybridation a été définie expérimentalement

<sup>b</sup> Ajout de mésappariements par rapport à la séquence nucléotidique de l'ADN cible

Les étalements et cultures liquides ont été effectués selon les conditions précisées dans le sous chapitre III-1 et en présence des antibiotiques adéquats, selon les concentrations indiquées dans le sous chapitre III-1. Du L(+) arabinose (Carl Roth, Lauterbourg, France) à  $10 \text{ g.L}^{-1}$  et de l'IPTG (Euromedex) à  $500 \text{ }\mu\text{g.mL}^{-1}$  ont été ajoutés dans le milieu lorsque précisé.

Les plasmides utilisés ou construits sont répertoriés dans le **Tableau 10**. Les kits Macherey Nalgen NucleoSpin Plasmid et NucleoBond PC100 ont été utilisés pour les extractions plasmidiques selon les recommandations du fournisseur. La qualité et la concentration de l'extrait plasmidique ont été vérifiées par dosage au spectrophotomètre ND-1000 (Nanodrop Technologies, Wilmington, United States of America). Le plasmide extrait a également été contrôlé par digestion enzymatique.

## 2.2 Manipulations enzymatiques de l'ADN

Les réactions d'amplification ont été réalisées comme décrit dans le sous chapitre III-1. Les amorces ont été définies sous oligo 6.1 et commandées chez InvitroGen (Cergy Pontoise, France). Les amorces et les conditions d'amplifications utilisées sont répertoriées dans le **Tableau 11**. Les réactions de digestion ont été effectuées avec les enzymes de restriction Fermentas (Mundolsheim, France) selon les recommandations du fournisseur. Les fragments ADN ont été traités lorsque précisé avec l'enzyme Klenow Fragments (Roche Diagnostics, Meylan, France) et l'enzyme Phosphatase Alcaline Shrimp SAP (Roche Diagnostics) selon les recommandations du fabricant. La taille des fragments ADN a été contrôlée par migration sur gel d'agarose en présence de marqueur de taille (1Kb+, InvitroGen, Cergy Pontoise, France) et purifiés à l'aide du kit Macherey Nalgen NucleoSpin Extract II (Hoerd, France), soit directement sur colonne, soit à partir d'un gel (**Annexe 3**). Les fragments d'ADN ont été quantifiés visuellement sur gel en présence de marqueur de masse (MassRuler, Fermentas). Les réactions de ligation ont été effectuées avec l'enzyme T4 DNA ligase (Roche Diagnostics) selon les recommandations du fabricant.

## 2.3 Test d'efficacité des constructions toxiques

Une colonie fraîche a été inoculée en milieu liquide contenant les antibiotiques adéquats. Après 12 h incubation à  $37^{\circ}\text{C}$  et 150 rpm, la culture bactérienne a été diluée en série au  $1/10^{\text{ème}}$  dans une solution de NaCl à  $8 \text{ g.L}^{-1}$  jusqu'à la dilution  $10^{-6}$ . Cent  $\mu\text{L}$  de chacune des dilutions ont été étalés sur boîte de Petri contenant du milieu nutritif gélosé contenant du chloramphénicol en présence ou absence d'inducteur(s) (arabinose et / ou IPTG). Après

**Tableau 12 : Taux d'échappement des gènes toxiques dans une construction génétique donnée**

Gènes Toxiques	Espèces	Inducteur	Construction génétique	Taux échappement <sup>a</sup>	Référence
<b>mazF</b>	<i>Escherichia coli</i> MC4100	Température 42°C	<b>Chromosome</b> $\lambda$ cl857, mazEF <b>Plasmide</b> $\lambda$ P <sub>L</sub> ::mazF	10E-3 10E-4	Aizeman <i>et al.</i> , 1996
	<i>Escherichia coli</i>	Absence Thymine Triméthoprim Sulfonamide	<b>Chromosome</b> P <sub>2</sub> P <sub>3</sub> mazE-mazF RelA	10E-2	Sat <i>et al.</i> , 2003
<b>sspC</b>	<i>Escherichia coli</i>	IPTG	<b>Plasmide</b> P <sub>lac</sub> ::sspC	10E-2	Setlow <i>et al.</i> , 1992
<b>ccdB</b>	<i>Escherichia coli</i>		<b>Plasmide</b> ccdA-ccdB-geneD		Couturier <i>et al.</i> , 1998
<b>nuc</b>	<i>Escherichia coli</i>	Température 42°C	<b>Chromosome</b> $\lambda$ cl857, <b>Plasmide</b> P <sub>L</sub> ::nuc	10E-5	Ahrenholtz <i>et al.</i> , 1994
<b>colE3</b>	<i>Escherichia coli</i> CC118		<b>Chromosome</b> lacI <sup>R</sup> , P <sub>trc</sub> P <sub>bla</sub> ::immE3 <b>Plasmide</b> lacI <sup>q</sup> , P <sub>lac</sub> ::colE3	10E-5	Diaz <i>et al.</i> , 1994
	<i>Pseudomonas putida</i>		<b>Chromosome</b> lacI <sup>R</sup> , P <sub>trc</sub> P <sub>bla</sub> ::immE3 <b>Plasmide</b> lacI <sup>R</sup> , P <sub>tac</sub> ::colE3	10E-4 10E-5	Diaz <i>et al.</i> , 1994
	<i>Rhizobium meliloti</i> (réceptrice)			10E-4	Diaz <i>et al.</i> , 1994
	<i>Agrobacterium tumefaciens</i> (réceptrice)			>10E-4	Diaz <i>et al.</i> , 1994
	<i>Comonas acidovorans</i> (réceptrice)			10E-4	Diaz <i>et al.</i> , 1994
	<i>Alcaligenes eutrophus</i> (receveuse)			>10E-5	Diaz <i>et al.</i> , 1994
	<i>Pseudomonas fluorescense</i> (réceptrice)			10E-6	Diaz <i>et al.</i> , 1994
	<i>Pseudomonas putida</i> KT2442 (pCC102)	Absence Benzoate IPTG	<b>Chromosome</b> P <sub>A1-03/04</sub> ::colE3 <b>Plasmide</b> xylS2, P <sub>m</sub> ::lacI	5.10E-4 10E-5	Munthali <i>et al.</i> , 1996
	<i>Escherichia coli</i> HB101	(IPTG)	<b>Chromosome</b> P <sub>trc</sub> ::immE3 <b>Plasmide</b> lacI <sup>R</sup> , P <sub>trc</sub> P <sub>c</sub> ::colE3	10E-5	Torres <i>et al.</i> , 2003
	<i>Escherichia coli</i> HB101	(IPTG)	<b>Chromosome</b> P <sub>trc</sub> ::immE3, P <sub>i</sub> ::ecoR1M <b>Plasmide</b> lacI <sup>R</sup> , P <sub>c</sub> ::colE3	10E-5	Torres <i>et al.</i> , 2003
<b>ecoR1R</b>	<i>Escherichia coli</i> (donatrice)	(IPTG)	<b>Chromosome</b> P <sub>trc</sub> P <sub>bla</sub> ::EcoR1M <b>Plasmide</b> lacI <sup>R</sup> P <sub>trc</sub> ::EcoR1R	10E-5	Torres <i>et al.</i> , 2000
	<i>Escherichia coli</i> (donatrice)	(IPTG)	<b>Chromosome</b> P <sub>trc</sub> ::immE3, P <sub>i</sub> ::ecoR1M <b>Plasmide</b> lacI <sup>R</sup> , P <sub>trc</sub> ::EcoR1R	10E-4	Torres <i>et al.</i> , 2003
	<i>Agrobacterium tumefaciens</i> (réceptrice)			10E-4	Torres <i>et al.</i> , 2000
	<i>Ralstonia eutropha</i> (réceptrice)			10E-4	Torres <i>et al.</i> , 2000
	<i>Pseudomonas putida</i> (réceptrice)			10E-4	Torres <i>et al.</i> , 2000

Gènes Toxiques	Espèces	Inducteur	Construction génétique	Taux échappement <sup>1</sup>	Référence
<b>stv</b>	<i>Pseudomonas putida</i>	Absence benzoate IPTG	<b>Plasmide</b> <i>xyIS2</i> , <i>P<sub>m</sub>::lacI</i> , <i>P<sub>10</sub>::stv</i> <i>P<sub>tet</sub>::lys</i> , <i>T7RNA::P<sub>trp</sub>-lac(tac)</i>	10E-4 10E-5	Szafranski <i>et al.</i> , 1997
			<b>Plasmide</b> <i>xyIS2</i> , <i>P<sub>m</sub>::lacI</i> , <i>stv::P<sub>10</sub></i> <i>P<sub>tet</sub>::lys</i> , <i>T7RNA::P<sub>trp</sub>-lac(tac)</i>	10E-5 10E-6	Szafranski <i>et al.</i> , 1997
			<b>Plasmide</b> <i>xyIS2</i> , <i>P<sub>m</sub>::lacI</i> , <i>stv::P<sub>10</sub></i> , <i>O<sub>lac</sub></i> <i>P<sub>tet</sub>::lys</i> , <i>T7RNA::P<sub>trp</sub>-lac(tac)</i>	10E-6 10E-7	Szafranski <i>et al.</i> , 1997
<b>sacB</b>	<i>Escherichia coli</i>	saccharose	<b>Chromosome</b> <i>nptI</i> - <i>sacR-B</i>	10E-5	Recorbet <i>et al.</i> , 1993
<b>hok</b>	<i>Escherichia coli</i> JM101, <i>lacI<sup>q</sup></i>	IPTG	<b>Plasmide</b> <i>P<sub>lac</sub>::hok</i>	10E-2	Bej <i>et al.</i> , 1988
<b>gef</b>	<i>Pseudomonas putida</i> KT2440 (pWW0, pCC102)	Absence benzoate, IPTG	<b>Chromosome</b> <i>P<sub>A1-04/03</sub>::gef</i> <b>Plasmide</b> <i>xyIS P<sub>m</sub>::lacI</i> , TOL(pWW0)	10E-5	Jensen <i>et al.</i> , 1993
	<i>Pseudomonas putida</i> EEZ29 (pWW0, pCC102)	Absence benzoate, IPTG	<b>Chromosome</b> <i>P<sub>lac</sub>::gef</i> <b>Plasmide</b> <i>xyIS P<sub>m</sub>::lacI</i> , TOL	10E-6 10E-7	Ronchel <i>et al.</i> , 1995
	<i>Pseudomonas putida</i> KT2440 (pWW0)	Absence benzoate, IPTG	<b>Chromosome</b> <i>lacI<sup>q</sup> P<sub>A1-04/03</sub>::gef</i>	10E-6	Jensen <i>et al.</i> , 1993
	<i>Pseudomonas putida</i> KT2440 (pWW0)	Absence benzoate, IPTG	<b>Chromosome</b> <i>P<sub>lac</sub>::gef</i> <i>xyIS</i> , <i>P<sub>m</sub>::lacI</i>	10E-8	Molina <i>et al.</i> , 1998
	<i>Pseudomonas putida</i> KT2440 V. <i>asd</i> (pWW0)	Absence benzoate, IPTG	<b>Chromosome</b> <i>P<sub>A1-03/04</sub>::gef</i> <i>xyIS</i> , <i>P<sub>m</sub>::lacI</i> <i>xyIS</i> , <i>P<sub>m</sub>::asd</i>	10E-9	Ronchel <i>et al.</i> , 2001
<b>gef-gef</b>	<i>Pseudomonas putida</i> KT2440 (pWW0)	Absence benzoate, IPTG	<b>Chromosome</b> <i>lacI<sup>q</sup></i> , <i>P<sub>A1-04/03</sub>::gef</i> <i>P<sub>A1-04/03</sub>::gef</i>	10E-8	Jensen <i>et al.</i> , 1993
<b>relF</b>	<i>Escherichia coli</i>	IPTG	<b>Chromosome</b> <i>lacI<sup>q</sup></i> <b>Plasmide</b> <i>P<sub>lacUV5</sub>::relF</i>	10E-5 10E-6	Knudsen <i>et al.</i> , 1991
<b>relF-relF</b>	<i>Escherichia coli</i>	IPTG	<b>Chromosome</b> <i>lacI<sup>q</sup></i> <b>Plasmide</b> <i>P<sub>A1-03/04</sub>::2x(relF)</i>	10E-7	Knudsen <i>et al.</i> , 1991
<b>colE3-EcoR1R</b>	<i>Escherichia coli</i>	IPTG	<b>Chromosome</b> <i>lacI<sup>q</sup></i> , <i>lacZ</i> <b>Plasmide</b> <i>2x(P<sub>A1-03/04</sub>::relF)</i>	10E-8	Knudsen <i>et al.</i> , 1995
	<i>Escherichia coli</i>	(IPTG)	<b>Chromosome</b> <i>P<sub>trc</sub>::immE3</i> , <i>P<sub>i</sub>::ecoR1M</i> <b>Plasmide</b> <i>lacI<sup>q</sup></i> , <i>P<sub>trc</sub>::EcoR1R</i> , <i>P<sub>c</sub>::colE3</i>	10E-6	Torres <i>et al.</i> , 2003
	<i>Escherichia coli</i>	(IPTG)	<b>Chromosome</b> <i>P<sub>trc</sub>::immE3</i> , <i>P<sub>i</sub>::ecoR1M</i> <b>Plasmide</b> <i>lacI<sup>q</sup></i> , <i>P<sub>trc</sub>::EcoR1R</i> , <i>P<sub>c</sub>::colE3</i>	10E-8	Torres <i>et al.</i> , 2003

<sup>1</sup> La méthode de calcul du taux d'échappement est variable suivant le contexte d'étude.

incubation 1 nuit à 37°C, les colonies (UFC) ont été dénombrées. Un taux d'échappement a été calculé en divisant le nombre d'UFC.mL<sup>-1</sup> se développant sur milieu avec inducteur par le nombre d'UFC.mL<sup>-1</sup> se développant sur milieu sans inducteur.

### **3 RESULTATS ET DISCUSSION**

#### **3.1 Choix des gènes toxiques pour la construction de la cassette toxique**

Le choix des gènes toxiques pour la construction de la cassette relève d'une étude de la littérature sur l'efficacité de différents gènes toxiques exploités ces dernières années dans un contexte d'étude donné (**Tableau 12**). Bien que la comparaison des efficacités des différents systèmes construits doive être effectuée avec prudence, certaines tendances peuvent néanmoins être dégagées.

Dans un système d'expression inductible simple constitué du gène toxique et du gène réprimant son expression en condition non inductrice, les gènes toxiques présentent une efficacité variable. Certains gènes semblent en effet moins efficaces, avec un taux d'échappement élevé. Le gène *sspC* dont l'action toxique cible l'ADN a un taux d'échappement de 10<sup>-2</sup>. Ce gène toxique aurait en effet tendance à induire de forts taux de mutations sous conditions de stress (Setlow *et al.*, 1992). L'expression du gène *ccdB*, codant une toxine qui bloque la gyrase et entraîne des lésions de l'ADN, a également des effets mutagènes importants (Couturier *et al.*, 1998).

D'autres gènes toxiques ciblant l'ADN montrent un taux d'échappement plus faible comme *EcoRIR* (Torres *et al.*, 2000). Ce type de gènes toxiques est potentiellement intéressant dans le cadre de l'endiguement génétique. Leur expression entraîne non seulement la mort de la cellule mais également la dégradation de son matériel génétique. Néanmoins, leur utilisation comme système de contre sélection positive est délicate dans le contexte de notre étude. Toute expression basale du gène en absence d'inducteur entraînerait la dégradation du matériel génétique, dont le plasmide portant les zones de recombinaison. Elle serait de plus susceptible de provoquer des mutations suite à l'induction du système SOS pour la réparation de l'ADN, comme pour *sspC* ou *ccdB*. Aussi, nous avons fait le choix d'éviter tout gène dont l'activité toxique cible l'ADN.

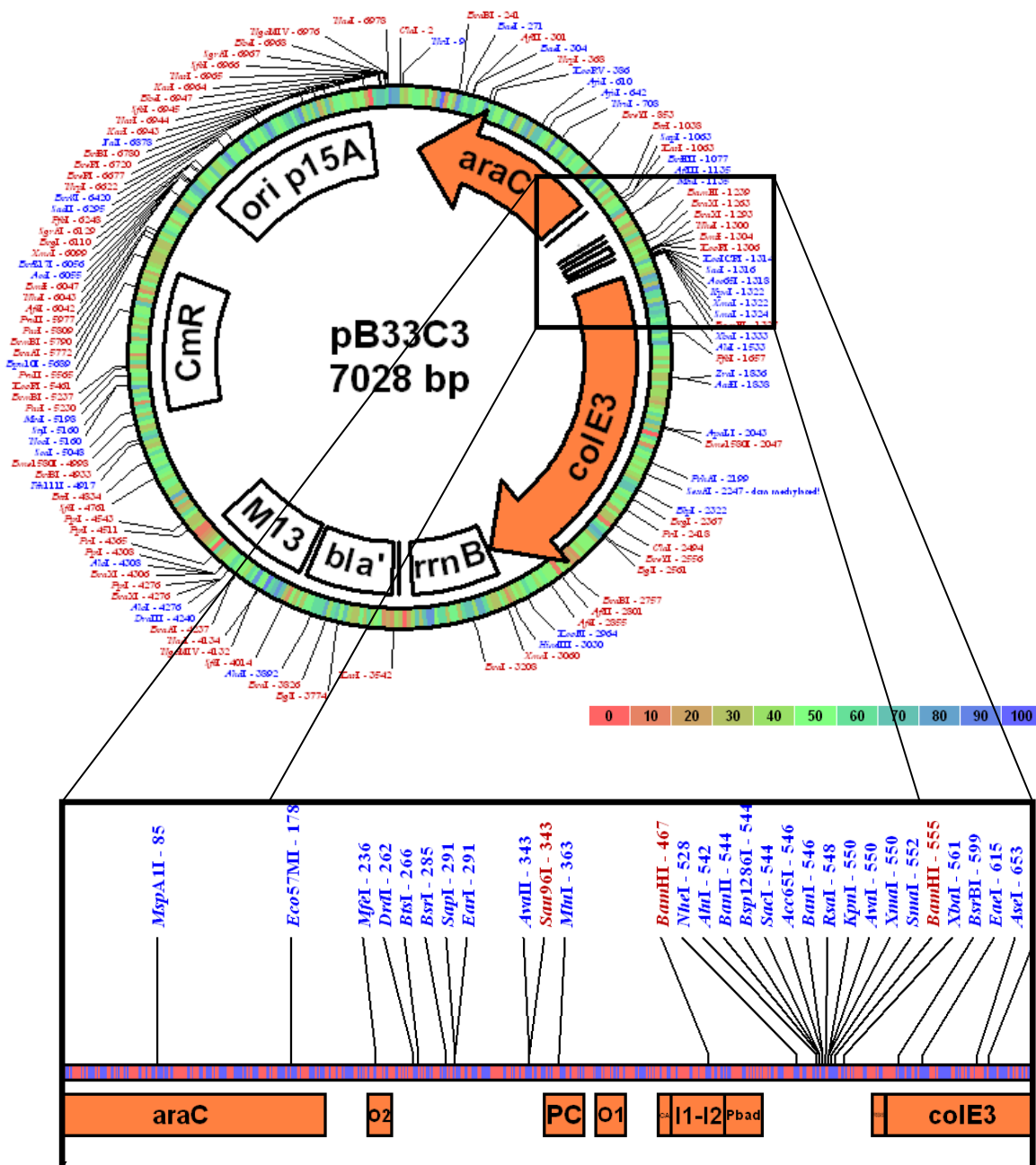
D'autres gènes toxiques ciblant la membrane, l'ARNr ou une autre composante essentielle au fonctionnement de la cellule ont été expérimentés, certains gènes présentant une forte efficacité notamment les gènes *relF*, *gef*, et *colE3*. Le mode d'action des gènes *gef* et *relF* laisse présager un spectre d'action large. Tous deux codent une porine, qui entraîne la déplétion du potentiel membranaire, aboutissant à la mort cellulaire (Molin *et al.*, 1993). Le gène *colE3*, codant une colicine à activité RNAsique, a été testé avec succès dans différentes souches environnementales. Son expression a entraîné la mort des bactéries testées (Diaz *et al.*, 1994). Les gènes *relF*, *gef* et *colE3* semblent donc être des gènes toxiques de choix pour la construction d'un système de contre sélection positive.

Un travail préliminaire effectué sur les gènes *relF* et *colE3* a permis de choisir un système régulateur adapté au contrôle de l'expression de chacun des deux gènes toxiques. L'expression du gène *relF* est ainsi sous le contrôle de la protéine LacI et le gène *colE3* sous le contrôle de la protéine AraC. Le contrôle de l'expression de *colE3* a notamment conduit au développement d'un nouveau système régulateur. Ce système fait intervenir l'expression constitutive de *immE3* à faible taux, pour inhiber l'activité toxique de la colicine exprimée de manière résiduelle. Un tel système annule tout effet nocif phénotypiquement visible sur la souche hôte, soit par l'aspect des colonies, soit par la vitesse de croissance de la souche hôte. Par ailleurs, la transformation normalisée par mole d'ADN d'un plasmide contenant *colE3* avoisine  $10^0$  grâce au système régulateur développé (sous chapitre III-1). Un tel système de contrôle d'expression laisse envisager un recouvrement possible d'une construction toxique par clonage successif du gène *colE3* puis du gène *relF* sur un même plasmide.

### **3.2 Contrôle de la souche réceptrice**

Dans la stratégie de construction de la cassette toxique inductible, les deux gènes toxiques sont clonés dans un vecteur similaire à celui qui sera utilisé pour la construction du système de contre sélection final. Ce choix permet de contrôler l'efficacité des gènes toxiques dans le vecteur final utilisé. Ce vecteur ne possédant pas le gène *lacI<sup>q</sup>*, celui-ci a été inséré dans le chromosome de la souche hôte (sous chapitre III-1) pour éviter d'augmenter la taille de la cassette toxique inductible. Seul le gène répresseur *araC* est apporté dans la cassette toxique inductible pour les raisons évoquées ci-après.





**Figure 11 : Carte de restriction de pB33C3**

Dessin réalisé avec le logiciel pDRAW32 d'après la séquence reconstruite de pB33C3. En orange : la cassette toxique, en bleu : les sites de restriction unique, en rouge : les sites de restriction double. Pc : promoteur de l'ORF *araC*, O1 et O2 : séquences opératrices, I1 et I2 : sites d'initiation de la transcription, CAP : site de fixation de la protéine CAP pour la répression catabolique.

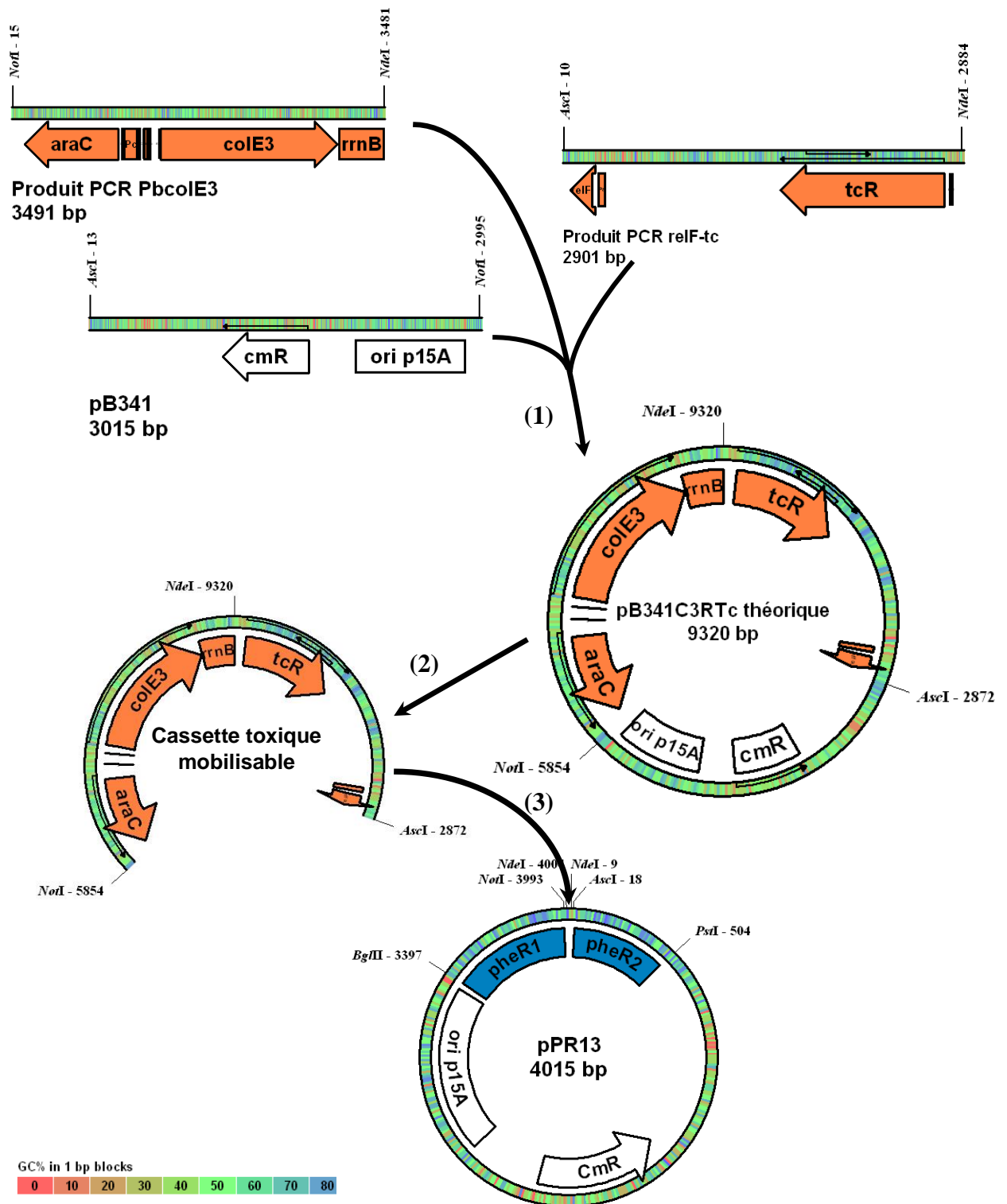
Le gène *lacI<sup>q</sup>* a donc été transféré dans le chromosome dans la souche *E. coli* 1661, via la cassette [P3::*immE3-kn'-lacI<sup>q</sup>*] selon le mode opératoire décrit dans le sous chapitre III-2. La fonctionnalité de *immE3* a auparavant été contrôlée, notamment par la capacité de la souche à acquérir par transformation un plasmide contenant le gène *colE3* (tests réalisés avec les plasmides pB33C3 et pPRE, voir sous chapitre III-1 et sous chapitre III-2). La fonctionnalité de *lacI<sup>q</sup>* a été vérifiée par transformation avec succès du plasmide pSK360 dans la souche *E. coli* 1661LIΔara. Ce plasmide porte deux gènes toxiques *relF*, dont l'expression est réprimée par LacI, mais ne porte pas le gène *lacI*. Aussi, la transformation de ce plasmide dans *E. coli* 1661LIΔara n'est permise que par l'expression du gène *lacI<sup>q</sup>* inséré dans le chromosome. La fréquence de transformation obtenue pour ce plasmide est faible, néanmoins des clones ont pu être isolés, sensibles à l'ajout d'IPTG en milieu liquide. Le gène *lacI<sup>q</sup>* inséré est donc fonctionnel.

### 3.3 Construction d'une cassette toxique inductible à deux gènes toxiques

La cassette toxique inductible doit pouvoir être facilement mobilisée pour son clonage entre les régions cibles dévolues à la recombinaison du système de contre sélection. Ceci implique le clonage des deux gènes toxiques de la cassette à proximité de façon à limiter la taille de la cassette. Toutes les séquences ADN non indispensables sont dans la mesure du possible éliminées. Par ailleurs, des sites de restriction sont ajoutés aux extrémités de la cassette.

#### Clonage de *relF* dans pB33C3

Le plasmide pB33C3 contenant *colE3* sous le contrôle de P<sub>bad</sub> avait été construit dans une étape précédente (sous chapitre III-1). D'après la carte de restriction plasmidique de pB33C3 (**Figure 11**), le site de restriction MluI a été identifié comme le plus adéquat pour y cloner la cassette [P<sub>A1-O3/O4</sub>::*relF-tc'*]. Après insertion de la cassette au niveau de ce site de restriction, le plasmide pB33C3RTc3G6 a pu être isolé. Ce dernier confère la résistance à la tétracycline, présente le profil de restriction attendu, et tue la cellule hôte en présence d'IPTG. La cassette [P<sub>A1-O3/O4</sub>::*relF-tc'*] a donc bien été insérée dans pB33C3, et le gène *relF* est fonctionnel. Le taux d'échappement lié à l'activité seule de *relF* est de 5.10<sup>-4</sup> à 5.10<sup>-5</sup>.



**Figure 12 : Stratégie de construction et de mobilisation de la cassette toxique**

(1) Digestion des trois produits PCR par SgsI (AscI), NotI et NdeI puis ligation pour obtenir pB341C2RTc (2) Digestion de pB341C3RTc par SgsI (AscI) et NotI pour mobiliser la cassette toxique et (3) l'insérer via les sites NotI et SgsI (AscI) dans un vecteur contenant les deux sites de recombinaison définis pour la capture de la diversité des gènes ciblés. En orange : la cassette toxique, en bleu : les sites de recombinaison.

Cependant l'ajout d'arabinose n'entraîne plus la mort de la cellule hôte, contrairement au vecteur pB33C3. La fonctionnalité du gène *colE3* aurait donc été perdue lors de l'étape de clonage. L'analyse de la séquence du plasmide pB33C3 indique que le site MluI utilisé pour le clonage est situé dans le promoteur du gène répresseur *araC* (**Figure 11**). Le clonage de la cassette dans son promoteur perturberait donc la répression du gène *colE3* d'une part, mais également de l'activation de sa transcription. La présence du gène immunité dans le chromosome de la souche hôte permet la stabilisation du plasmide contenant *colE3*, cependant l'absence d'activation expliquerait la survie des cellules hôtes observée en présence de l'inducteur arabinose.

Le clonage successif des deux gènes toxiques *relF* et *colE3* sur un même plasmide est techniquement réalisable. Des contraintes moléculaires ne nous ont pas permis d'obtenir une construction fonctionnelle. Les sites de restriction pour le clonage de la cassette [ $P_{A1-O3/O4}::relF-tc'$ ] à proximité de *colE3* dans pB33C3 étant en nombre limité, nous avons décidé de cloner ces deux gènes *relF* et *colE3* sur un nouveau vecteur.

#### Clonage des gènes *relF* et *colE3* dans pB341

Pour cloner les gènes *colE3* et *relF* dans une cassette toxique finale aisément mobilisable, la stratégie de clonage suivante a été appliquée : le vecteur et les deux gènes toxiques ont été amplifiés par des couples d'amorces portant des sites de restriction afin de réaliser un clonage orienté en une seule étape. Cette stratégie de clonage a déjà été réalisée avec succès pour le clonage de deux sites de recombinaison dans le plasmide pB342 (sous chapitre III-2). La zone de pBAD33 comprise entre  $P_{bad}$  et l'ORF *araC* est composée de nombreux opérateurs du promoteur du gène *araC* ainsi que des opérateurs de  $P_{bad}$ , rendant délicate la séparation entre *araC* et  $P_{bad}$ . Aussi le gène *araC* est conjointement cloné avec  $P_{bad}::colE3$  dans la cassette toxique inductible.

Le couple d'amorces Bad33ascIFw3 / Bad33NotIRv2 a été défini d'après la séquence reconstruite de pBAD33 afin d'amplifier un fragment d'ADN qui sera utilisé comme vecteur (**Figure 12**). La cassette [ $P_{A1-O3/O4}::relF-tc'$ ] est amplifiée à partir du plasmide pSK360 par le couple d'amorces ParelFTcAscIFw2 / ParelFTcNdeIRv5 et la cassette [*araC*- $P_{bad}::colE3-rnB$ ] à partir de pB33C3 par le couple d'amorces PbcolE3notIFw5 / PbcolE3ndeIRv5. Les sites de restriction SgsI, NdeI et NotI ont été ajoutés sur les amorces (par modification de nucléotides) pour le clonage orienté de chacun des produits PCR (**Figure 12**). La cassette toxique finale constituée de [*araC*- $P_{bad}::colE3-rnB$ ] et [ $P_{A1-O3/O4}::relF-tc'$ ] pourra ainsi être mobilisée par digestion avec les enzymes SgsI et NotI. L'amplification PCR, suivie d'une

digestion puis d'un clonage n'a cependant pas permis l'obtention de transformants en dépit de la répétition des tentatives.

Dans une stratégie alternative, chacun des deux gènes toxiques a été cloné successivement sur le plasmide pB341. La cassette [*araC*-P<sub>bad</sub>::*colE3*-rrnB] a été amplifiée par le même couple d'amorces PbcolE3notIFw5 / PbcolE3ndeIRv5, mais seul le site de restriction NotI présent sur l'amorce sens est utilisé pour le clonage de la cassette dans pB341. Le plasmide pB341C3 obtenu montre une activité toxique en présence d'arabinose. La cassette [P<sub>A1-O3/O4</sub>::*relF*-*tc'*] est amplifiée par le couple d'amorces ParelFTcAscIFw1 / ParelFTcascIRv2 à partir de pSK360 pour son clonage dans pB341C3. Les plasmides pB341C3RTc1 obtenus confèrent la résistance à la tétracycline, cependant l'ajout d'IPTG n'entraîne pas la mort de la cellule hôte. Le profil de digestion de pB341RTc1 indique la présence d'un insert inférieur à la taille attendue. L'analyse détaillée du couple d'amorces utilisées pour l'amplification de [P<sub>A1-O3/O4</sub>::*relF*-*tc'*] indique une hybridation annexe sur un autre emplacement de pSK360 pouvant expliquer le clonage d'un fragment contenant *tc'* sans *relF*. Le sous-produit PCR pourtant non détecté sur gel a été cloné préférentiellement au produit PCR majoritaire contenant *relF*.

L'obtention d'une construction comportant les gènes *relF* et *colE3* fonctionnels sur un même plasmide a nécessité le clonage de la cassette [P<sub>A1-O3/O4</sub>::*relF*-*tc'*] dans le plasmide pB341, la digestion de pB341RTc réalisée par SgsI pour ré-extraire la cassette et la cloner dans pB341C3. La souche utilisée pour toutes les réactions de clonage ne présentant pas d'inactivation du gène *hsdR* (sous chapitre III-2), les produits PCR pénétrant dans la cellule lors de l'étape de transformation sont donc dégradés et non intégrés. Ceci expliquerait les échecs de clonage rencontrés lors du clonage simultané des deux gènes toxiques dans le plasmide pB341. La souche *E. coli* DH5αZ1 utilisée lors de cette dernière opération de clonage est quant à elle mutée dans son système de restriction. Les ADN non méthylés entrant dans la cellule ne sont donc pas dégradés, et des clones présentant les génotypes plasmidiques et les phénotypes attendus ont pu être isolés, confirmant ainsi la validité de l'hypothèse. Le système de modification de la souche *E. coli* DH5αZ1 n'est quant à lui pas muté. Le plasmide pB341RTc extrait de cette souche est donc méthylé, par conséquent la cassette [P<sub>A1-O3/O4</sub>::*relF*-*tc'*] issue de ce plasmide aussi. Ceci explique le recouvrement de la construction toxique attendue seulement après clonage de cette cassette issue de *E. coli* DH5αZ1. L'obtention de transformants dans la souche *E. coli* 1661LIΔara possédant pB341C3 malgré la possible restriction des produits PCR peut être expliquée par une saturation du système de

restriction / modification via la pénétration d'une grande quantité d'ADN (Nielsen *et al.*, 1998).

Le plasmide pB341C3RTc2 obtenu comporte une cassette mobilisable constituée de 2 gènes toxiques inductibles et entraîne la mort des cellules hôtes après induction par l'IPTG, ainsi que par l'arabinose. Les gènes *relF* et *colE3* fonctionnels ont pu être clonés avec succès sur un même plasmide.

### 3.4 Efficacité de la cassette toxique inductible

Le test d'efficacité de la construction toxique portée par pB342C3RTc2 dans *E. coli* 1661LIΔara par induction simultanée de l'expression des deux gènes toxiques indique un taux d'échappement de  $10^{-7}$ . Le clonage de deux gènes toxiques sur le même plasmide a permis un gain d'efficacité, mais ce gain est inférieur à ce qui était attendu. La construction est seulement 10 fois plus efficace alors qu'un taux d'échappement de  $10^{-8}$  à  $10^{-11}$  était attendu. Le test d'efficacité réalisé par induction de chacun des gènes toxiques de manière individuelle indique un taux d'échappement de  $10^{-6}$  pour le gène *colE3* et de seulement  $10^{-3}$  pour le gène *relF*. La même cassette clonée dans pMMB190 permettait un taux d'échappement de  $10^{-5}$  dans pMMB190 (sous chapitre III-1). Cette perte d'activité peut être liée soit au vecteur, soit à la souche hôte utilisée.

Le vecteur possède une origine de réplication différente de pMMB190 et ne porte pas le gène *lacI*. Les résultats obtenus pour pB33C3RTc3G6 indiquent une perte d'efficacité du gène *relF* de 5 à 50 fois lorsque la construction est clonée sur un dérivé de pBAD33 par rapport à pMMB190. Pour le plasmide pB341C3RTc2, l'efficacité du gène *relF* est diminuée d'un facteur 100 par rapport à pMRTC20. Le clonage du gène *lacI<sup>q</sup>* au niveau chromosomique entraîne une plus faible expression de LacI comparativement au même gène cloné sur un plasmide. La quantité de protéine produite est en effet proportionnelle au nombre de copies du vecteur. La répression pourrait être insuffisante, entraînant une expression basale du gène *relF*. Cette expression basale conduirait à une réponse hormétique de la cellule, expliquant ainsi la perte d'efficacité observée.

Cette hypothèse n'est cependant pas tout à fait valide. Le gène *lacI<sup>q</sup>* lorsqu'il est exprimé entraîne en effet la production d'une quantité de protéines suffisante pour permettre la répression des gènes sous le contrôle du promoteur  $P_{lac}$  présent sur des vecteurs à faible nombre de copies ou à nombre de copies intermédiaire (Glascok & Weickert, 1998). Le problème dans ce cas ne viendrait donc pas du nombre de copies du vecteur, mais du gène *lacI<sup>q</sup>* inséré dans le chromosome de *E. coli* 1661LIΔara. L'expression de la protéine LacI a pu

être vérifiée phénotypiquement, la quantité de protéines exprimées n'a cependant pas été vérifiée, et aucun contrôle génotypique n'a encore été réalisé. Au cours des étapes d'intégration chromosomique, de transduction, et de modification chromosomique de la souche réceptrice, des événements ont pu conduire à la mutation du gène *lacI<sup>q</sup>* inséré, et consécutivement à la production d'une plus faible quantité de protéines que celle attendue.

Des analyses complémentaires seront nécessaires afin de détecter l'origine de cette perte d'efficacité du gène *relF*. Un test d'efficacité du plasmide pB341RTc contenant le gène *relF* seul dans *E. coli* 1661LIΔara d'une part et dans une autre souche contenant un *lacI<sup>q</sup>* chromosomique fonctionnel d'autre part, validerait ou infirmerait cette dernière hypothèse. Elle permettrait par ailleurs de vérifier si la présence du gène *colE3* sur le même plasmide a un impact ou non sur l'efficacité du gène *relF*.

Très récemment, une cassette toxique inductible contenant deux gènes toxiques différents a également été construite. Le taux d'échappement de la construction est de  $10^{-8}$  (Harms *et al.*, 2007). Ces derniers ont utilisé le gène *sacB* et le gène *hok*. Le gène *hok* présente un mode d'action similaire au gène *relF* (Molin *et al.*, 1993). Le gène *sacB* exprime une enzyme responsable d'une accumulation de levane dans le périplasme de la cellule hôte, menant à la lyse cellulaire (Recorbet *et al.*, 1993). L'ajout de saccharose étant nécessaire à l'activité toxique de *sacB*, les problèmes d'expression basale sont évités. Cependant, un examen de la littérature indique une efficacité de *sacB* équivalente à inférieure à celle de *relF* ou *colE3* (**Tableau 12**). Dans la construction toxique de ces auteurs (Harms *et al.*, 2007), le gène *sacB* a un taux d'échappement de  $10^{-6}$ , et le gène *hok* de  $10^{-5}$ . Les taux d'échappement de chacun des gènes sont équivalents à ceux obtenus pour *colE3* et *relF*. Un point cependant intéressant à noter est la localisation chromosomique de la construction toxique dans la souche bactérienne de ces auteurs (Harms *et al.*, 2007). Le passage d'une construction toxique d'un plasmide au chromosome permet généralement d'accroître son efficacité d'un ordre de grandeur (Molina *et al.*, 1998; Ronchel *et al.*, 1995). Les taux d'échappement que nous avons actuellement atteints sont observés pour des constructions plasmidiques. En réglant d'une part les problèmes d'efficacité liés au gène *relF* et en diminuant le nombre de copies du vecteur portant la cassette toxique, nous pouvons espérer augmenter significativement l'efficacité du système létal. L'utilisation d'un vecteur à une seule copie par cellule nécessite cependant certaines vérifications, notamment l'efficacité de *colE3* à tuer la cellule hôte en condition inductrice. Le système de régulation pour l'expression inductible de *colE3* ayant été développé pour un nombre moyen de copies, son fonctionnement sur un plasmide à faible nombre de copies doit être vérifié.

## 4 **CONCLUSIONS**

La construction d'une cassette toxique constituée de deux gènes toxiques différents, chacun sous le contrôle de systèmes régulateurs indépendants, a nécessité le déploiement de stratégies de clonage afin que la cassette toxique construite soit mobilisable, que les deux gènes toxiques soient fonctionnels, et que l'efficacité de la cassette toxique soit maximale.

Une cassette toxique inductible constituée des gènes *relF* et *colE3*, régulés respectivement par les protéines LacI et AraC, a ainsi été construite. L'efficacité de cette cassette toxique est supérieure à l'efficacité de toutes les constructions toxiques réalisées dans ce chapitre III. Alors que le taux d'échappement avec des cassettes constituées uniquement du gène *relF* ou du gène *colE3* est de  $10^{-5}$  à  $10^{-6}$ , le taux d'échappement de la nouvelle cassette construite atteint  $10^{-7}$ . La combinaison de deux gènes toxiques différents nous a ainsi permis d'augmenter l'efficacité tueuse de la cassette.

Un taux d'échappement inférieur à  $10^{-7}$  était cependant attendu. Des analyses complémentaires ont révélé que le gène *relF* présente une efficacité moindre dans la cassette toxique inductible construite dans ce sous chapitre par rapport à d'autres constructions réalisées. La compréhension de cette baisse d'efficacité et le rétablissement de l'efficacité du gène *relF* dans la cassette à deux gènes toxiques nous permettra d'atteindre et peut être même distancer les taux d'échappement actuellement reportés dans la littérature.

Cette optimisation de l'efficacité de la cassette construite est d'autant plus nécessaire si des événements rares sont recherchés. L'utilisation de cette cassette dans le système de contre sélection permettra la détection des souches recombinées à une fréquence de  $10^{-7}$ , si aucune zone d'homologie susceptible d'altérer l'efficacité du système n'est présente. Ce seuil de détection n'est cependant pas encore suffisamment bas si les événements de capture ont lieu à une fréquence inférieure.



## **SOUS CHAPITRE III-4 :**

### **PROSPECTION DES POTENTIELLES SOUCHES RECEPTRICES POUR ACCEDER A L'ADN DE LA MICROFLORE TELLURIQUE**

#### **1 INTRODUCTION**

Plusieurs aspects de la nouvelle approche métagénomique proposée dans cette thèse ont été présentés dans les précédents sous chapitres. Des premières études ont été réalisées sur la recombinaison de l'ADN de la microflore tellurique au niveau du système de contre sélection et sur le développement d'une cassette toxique pour sélectionner efficacement les souches réceptrices ayant capturé l'ADN de la microflore indigène. Dans ce dernier sous chapitre est traité un autre point clé de cette nouvelle approche : les voies d'entrée de l'ADN de la microflore tellurique à l'intérieur de la souche réceptrice.

Les bactéries sont en effet capables d'acquérir l'information génétique d'autres bactéries selon plusieurs voies : la transduction (transfert d'ADN entre deux bactéries via un virus), la conjugaison (transfert d'ADN entre deux bactéries via un contact physique), la transformation ou l'électrotransformation (acquisition d'ADN extracellulaire nu). Ces transferts d'ADN impliquent la rencontre entre bactéries, entre bactérie et ADN, ou entre bactéries et phages. Or, le sol est un environnement structuré, pouvant limiter la rencontre entre ces différentes entités (Dechesne *et al.*, 2005) suivant leur localisation dans les micro ou macroagrégats du sol (Mummey *et al.*, 2006; Mummey & Stahl, 2004) et les stratégies de distribution / colonisation dans le sol (Dechesne *et al.*, 2003; Grundmann *et al.*, 2001; Pallud *et al.*, 2004). Des événements de transferts d'ADN ont toutefois pu être détectés dans le sol à des fréquences de l'ordre de  $10^{-5}$  par transformation naturelle (Nielsen *et al.*, 1997a),  $10^{-4}$  à  $10^{-5}$  par électrotransformation (Ceremonie *et al.*, 2004) ou  $10^{-4}$  par conjugaison (Richaume *et al.*, 1989). Des événements de transferts d'ADN par transduction y ont également été détectés (Zeph *et al.*, 1988).

Chacun de ces modes de transferts d'ADN requière certaines caractéristiques qui permettront l'acquisition d'ADN. En fonction de ces caractéristiques, une souche réceptrice adaptée pour la capture de l'ADN de la microflore tellurique doit être choisie.

La transduction nécessite l'intervention d'un phage capable d'infecter la souche donatrice, de récupérer une partie de l'information génétique, puis de réinfecter une souche receptrice dans laquelle sera intégré l'ADN de la souche donatrice. Un tel transfert serait permis par des phages tempérés, qui ont un cycle de type lysogénique. Ces derniers s'intègrent dans le génome bactérien (prophage) et n'entrent dans une phase lytique qu'après induction (Marsh & Wellington, 1994; Weinbauer, 2004). Les phages peuvent alors transférer soit uniquement les gènes bactériens situés à proximité du site d'intégration du prophage (transduction spécialisée), soit n'importe quel gène bactérien (transduction généralisée) (Davison, 1999). Les transferts d'ADN peuvent être plus ou moins limités entre certaines espèces bactériennes suivant le spectre d'hôte du phage. Les études actuelles semblent montrer que les phages ont généralement un spectre d'hôte étroit (Weinbauer, 2004). Néanmoins peu de phages bactériens ont pour l'instant pu être isolés parmi la grande diversité révélée par les approches métagénomiques (Edwards & Rohwer, 2005). Par ailleurs, le spectre d'hôtes des phages actuellement isolés n'a pu être défini qu'en fonction du faible pourcentage des bactéries à ce jour cultivées *in vitro* (Amann *et al.*, 1995).

Lors du processus de conjugaison, l'information génétique est transférée d'une souche donatrice vers une souche receptrice. Ce processus nécessite la présence chez la souche donatrice de gènes particuliers pour permettre le contact physique entre les deux bactéries, puis pour mobiliser l'ADN vers la souche receptrice. Un transfert de plasmide par conjugaison a déjà été observé de la microflore indigène vers une souche receptrice modifiée (Lilley & Bailey, 1997). Néanmoins peu d'informations ont pour l'instant été rassemblées sur la présence et la distribution de plasmides conjugatifs dans la microflore indigène.

Par transformation naturelle, la bactérie internalise l'ADN présent dans son environnement extracellulaire. L'acquisition de cet ADN n'a lieu que si la bactérie est dans un état physiologique particulier dit de compétence. Cette compétence peut se développer à différents stades de la croissance bactérienne (Lorenz & Wackernagel, 1994), généralement lorsque la cellule est dans un état énergétique élevé (Palmen & Hellingwerf, 1997). Plus de 40 espèces bactériennes, d'au moins 24 genres bactériens sont actuellement décrites comme naturellement transformables *in vitro* (Lorenz & Wackernagel, 1994; Palmen & Hellingwerf, 1997). Des événements de transferts d'ADN par transformation naturelle dans la matrice sol ont été observés chez différentes espèces bactériennes (Demaneche *et al.*, 2001c; Nielsen *et al.*, 1997b; Nielsen *et al.*, 2000). Néanmoins, les bactéries introduites dans l'environnement sol perdent rapidement leur état de compétence *in situ* (Nielsen *et al.*, 1997b). Celui ci peut

être prolongé par l'ajout de nutriments dans l'inoculum bactérien (Nielsen *et al.*, 1997b) ou induite *in situ* par l'ajout de nutriments (Nielsen *et al.*, 1997a)

L'électrotransformation naturelle est une voie d'acquisition d'ADN récemment mise en évidence dans l'environnement sol (Demaneche *et al.*, 2001a). Cette variante de la transformation naturelle permet l'entrée de l'ADN dans la cellule suite à un choc électrique. Un champ électrique intense entraîne des réarrangements structuraux de la membrane, notamment des pores aqueux temporaires, donnant lieu à une perméabilisation de la membrane (Weaver, 1995). De nombreuses espèces bactériennes sont capables d'acquérir de l'ADN libre *in vitro* par électrotransformation (Chen *et al.*, 2006; Lefrançois & Sicard, 1997; Lofblom *et al.*, 2007; van der Rest *et al.*, 1999). De plus, cette voie de transformation ne requiert pas un stade de compétence particulier (Ceremonie *et al.*, 2004).

Dans ce sous chapitre, nous nous sommes plus particulièrement intéressés à ces deux dernières voies d'acquisition de l'ADN. Alors que la conjugaison est dépendante de la microflore tellurique et que la transduction implique l'emploi d'un phage en plus d'une souche réceptrice, l'acquisition d'ADN par (électro)-transformation naturelle est surtout dépendante de la souche réceptrice. Ce sous chapitre présente les travaux initiés pour choisir une espèce bactérienne qui pourrait ultérieurement être utilisée en tant que souche réceptrice. Cette souche réceptrice doit en effet posséder certaines caractéristiques pour optimiser la probabilité de capture de l'ADN de la microflore tellurique : elle doit acquérir l'ADN à très haute fréquence, être recombinogène, être sensible à l'expression de la cassette inductible, mais également être cultivable pour son ré-isolément après accomplissement de sa fonction dans le sol. Sa capacité à coloniser l'environnement sol est également un facteur à prendre en compte pour l'accès à l'information génétique présente dans les différentes fractions du sol. Alors que de nombreuses bactéries présentes dans le sol sont naturellement transformables (Lorenz & Wackernagel, 1994), seulement deux souches isolées du sol ont à ce jour été montrées électrotransformables (Ceremonie *et al.*, 2004). Une méthode pour isoler des bactéries électrotransformables à partir de la matrice sol a été initiée. Des tests pour leur sensibilité à un gène toxique ont également été conduits.

## 2 MATERIELS ET METHODES

### 2.1 Extraction des cellules du sol et purification sur gradient de Nycodenz

Les sols prélevés à Montrond et à Châteaurenard (Bertrand *et al.*, 2005) ont été utilisés pour l'extraction des cellules du sol. L'équivalent de 125 g de sol sec a été dispersé dans 200 mL d'hexamétaphosphate de sodium à 2 g.L<sup>-1</sup>, en présence de billes de verre, avec agitation pendant 1 h. Les particules grossières ont été sédimentées par centrifugation à 707 g pendant 15 min à 10°C. Le surnageant a été récupéré, filtré sur gaze, puis centrifugé à 7 408 g pendant 20 min à 10°C. Le surnageant a été éliminé, et les cellules sédimentées ont été remises en suspension dans 20 mL de NaCl à 8 g.L<sup>-1</sup> pré refroidi à 4°C. Dans un tube d'ultra centrifugation contenant 10 mL de solution de Nycodenz (Axis-Shield, Oslo, Norvège) à 0,8 g.mL<sup>-1</sup> pré refroidi à 4°C, 25 mL de la suspension bactérienne ont été délicatement déposés en surface. Le tout a été centrifugé à 9 000 rpm 40 min 10°C dans l'ultracentrifugeuse Kontron TGA-50 (rotor TST 2838 Swing out). L'anneau cellulaire interphasique a été prélevé, puis lavé par une succession de centrifugation à 4 000 g 10 min 4°C, élimination du surnageant et remise en suspension des cellules dans 20 mL, 1 mL, puis 0,5 mL d'eau ultra pure. Les cellules ont été stockées à 4°C.

### 2.2 Milieux, bactéries et plasmides

Les souches *Acinetobacter baylyi* BD413 (anciennement nommé *A. calcoaceticus* BD413) (Juni & Janik, 1969; Vanechoutte *et al.*, 2006) et *Pseudomonas fluorescens* N3 (Ceremonie *et al.*, 2004) stockées en glycérol 25 % à -80°C ont été étalées sur milieu LB contenant de l'agar à 10 g.L<sup>-1</sup>, les cultures liquides ont été effectuées dans du milieu LB seul avec agitation. Les cellules ont été incubées à 28°C pendant 24 à 48 h. Les cellules extraites du sol ont été étalées sur milieu TSA (Tryptic Soy Agar) puis incubées à 28°C pendant 24 à 72h00. Les antibiotiques adéquats ont été ajoutés dans le milieu selon les concentrations finales précisées dans le sous chapitre III-I

Les plasmides pSM1885 (*gfp*, *sm<sup>r</sup>*, *tc<sup>r</sup>*, ori incQ) (Richter *et al.*, 2003) et pFEB98 (2x (P<sub>A1-03/04</sub> ::*relF*), *amp<sup>r</sup>*, *tc<sup>r</sup>*, ori incQ) (sous chapitre III-1) ont été extraits de souches *E. coli* DH5α avec l'aide des kits commerciaux comme précédemment indiqué. L'extraction plasmidique à partir des souches *A. baylyi* et *P. fluorescens* a été effectuée par lyse alcaline, extraction des protéines au phénol / chloroforme et purification de l'ADN par précipitation isopropanolique.

La qualité et la concentration de l'extrait plasmidique ont été vérifiées par dosage au spectrophotomètre ND-1000 (Nanodrop Technologies, Wilmington, United States of America). Les plasmides extraits ont également été contrôlés par digestion enzymatique.

### 2.3 Préparation des cellules pour l'électrotransformation

Les cellules extraites du sol stockées à 4°C ont été utilisées sans traitement ultérieur pour l'électroporation. *A. baylyi* BD413 et *P. fluorescens* N3 ont été rendues électrocompétentes selon le mode opératoire suivant. Une colonie isolée fraîche a été inoculée en milieu liquide en présence de(s) pression(s) de sélection adéquate(s). La préculture a été incubée avec agitation orbitaire de 150 rpm jusqu'à phase stationnaire, puis a été diluée au 1/100<sup>ème</sup> dans du milieu LBm contenant le(s) antibiotique(s) adéquat(s). La culture a été incubée à 150 rpm jusqu'à une densité optique (DO) mesurée à 600nm de 0,6 maximum. A partir de cette étape, les cellules bactériennes ont été maintenues à une température maximale de 4°C sur glace ou dans des appareils réfrigérants. La culture bactérienne a été centrifugée à 4 000 g 10 minutes à 4°C, le surnageant a été éliminé, et les cellules sédimentées ont été lavées dans du glycérol 10 % pré refroidi à 4°C. Les étapes de centrifugation / lavage ont été effectuées 4 fois, avec un volume de glycérol 10 % froid de 1, 1/2, 1/5<sup>ème</sup> puis 1/50<sup>ème</sup> du volume initial de la culture bactérienne. Les cellules ont été reprises dans un volume final de glycérol froid 10 % de 1/100<sup>ème</sup> du volume initial, puis aliquotées et stockées à -80°C.

### 2.4 Electrotransformation

Dans une cuve d'électroporation de 2 mm (BTX, MP Biomedicals, Illkirch, France) pré-refroidie à 4°C, 50 µL de cellules puis une solution de plasmide ajusté à 5 µL avec de l'eau stérile ont été déposés. Un choc de 2,5 kV.cm<sup>-1</sup>, 200 Ω et 25 µF a été appliqué avec GenepulserII (Bio-Rad, Marne La Coquette, France). Après choc électrique du milieu LBm a été ajouté pour QSP 1 mL, puis les cellules ont été incubées en présence de DNAase 2 h à 28°C avec agitation douce (<150rpm). La suspension bactérienne a été diluée en série au 1/10<sup>ème</sup> dans une solution de chlorure de sodium (NaCl) à 8 g.L<sup>-1</sup>. Les dilutions 10<sup>0</sup>, 10<sup>-1</sup>, 10<sup>-2</sup> ont été étalées sur milieu gélosé sélectif adéquat pour dénombrer les transformants, les dilutions 10<sup>-4</sup>, 10<sup>-5</sup> et 10<sup>-6</sup> ont été étalées sur milieu gélosé non sélectif pour dénombrer les cellules réceptrices. Une fréquence de transformation a été calculée.

## 2.5 Dénombrements des électrotransformants sur lames

Cent  $\mu\text{L}$  d'une dilution d'électrotransformants et 900  $\mu\text{L}$  de milieu TSA en surfusion, contenant les antibiotiques adéquats, ont été déposés sur une lame de microscope, puis mélangés directement sur lame. Une lamelle de microscope a été immédiatement déposée sur le mélange. Après solidification de l'agar, la lamelle a été enlevée, la lame a été placée dans une boîte de Petri stérile contenant un papier stérile imbibé d'eau stérile.

La boîte a été incubée à 28°C 24 à 72 h. Les électro transformants ont été visualisés sous filtre GFP avec le microscope AxioScope Zeiss, équipé du logiciel AxioVision V4.3 et les photos effectuées avec AxioCam MRc5 (Zeiss, Oberkochen, Allemagne).

## 2.6 Test d'efficacité des constructions toxiques

Une colonie fraîche a été inoculée en milieu liquide contenant les antibiotiques adéquats. Après incubation à 150 rpm, la culture bactérienne a été diluée en série au  $1/10^{\text{ème}}$  en série dans une solution de NaCl à 8  $\text{g.L}^{-1}$  jusqu'à la dilution  $10^{-6}$ . Cent  $\mu\text{L}$  de chacune des dilutions ont été étalés sur boîte de Petri contenant du milieu nutritif gélosé contenant de l'ampicilline à 100  $\mu\text{g.mL}^{-1}$  et de la tétracycline à 20  $\mu\text{g.mL}^{-1}$  (en concentration finale) en présence ou absence d'IPTG. Après incubation 1 nuit, les colonies (UFC) ont été dénombrées, et un taux d'échappement calculé.

# 3 RESULTATS ET DISCUSSION

## 3.1 Développement d'une méthode pour la détection de souches électrotransformables cultivables

Pour détecter des bactéries telluriques capables d'acquérir de l'ADN par électrotransformation, des cellules extraites du sol sont électroporées en présence de pSM1885. Ce plasmide possède deux gènes de résistance aux antibiotiques streptomycine ( $sm^r$ ) et tétracycline ( $tc^r$ ) pour sélectionner les bactéries ayant acquis le plasmide. Néanmoins, les souches environnementales peuvent présenter la résistance à de multiples antibiotiques (Demaneche *et al.*, 2001c). Pour différencier les vrais électrotransformants des bactéries naturellement résistantes, le gène marqueur *gfp* est utilisé. Ce gène porté par le plasmide pSM1885 exprime constitutivement la protéine GFP (*green fluorescent protein*) qui, sous la longueur d'onde d'excitation  $\lambda_{\text{excitation}} = 488 \text{ nm}$ , émet un rayonnement de longueur d'onde  $\lambda_{\text{émission}} = 510 \text{ nm}$ . Les bactéries produisant la protéine GFP apparaissent alors vertes.

**Tableau 13 :** Comparaison des estimations des densités bactérienne d'une culture bactérienne suivant la méthode de dénombrement utilisée

Dénombrement	UFC.mL <sup>-1</sup> Mode1 <sup>a</sup>	UFC.mL <sup>-1</sup> Mode2 <sup>b</sup>	Ecart type
<b>1</b>	6,01E+08	5,20E+08	5,73E+07
<b>2</b>	5,83E+08	5,80E+08	2,12E+06
<b>3</b>	1,37E+09	2,67E+09	9,19E+08

<sup>a</sup> Mode 1 : dénombrement suite à l'étalement de différentes dilution d'une suspension bactérienne sur boîte de pétri contenant un milieu nutritif solide

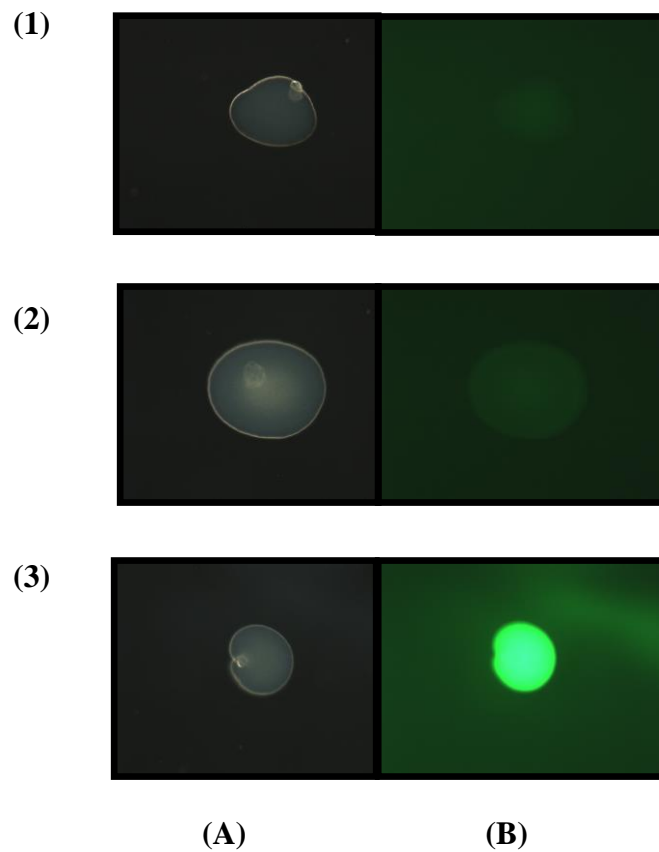
<sup>b</sup> Mode 2 : dénombrement suite au mélange d'une dilution donnée d'une suspension bactérienne avec un milieu nutritif en surfusion directement sur lame de microscope.

L'utilisation de boîtes de Petri pour étaler les transformants est cependant peu adaptée si un criblage ultérieur des colonies vertes est réalisé sous microscope à fluorescence. Chacune des colonies pourrait être visualisée entre lame et lamelle au microscope, mais le criblage serait long et fastidieux. Aussi nous avons expérimenté une nouvelle méthode d'étalement des bactéries, non plus sur boîte de Petri, mais directement sur lame de microscope. Un milieu gélosé serait déposé sur cette lame, dans lequel pourraient croître des colonies bactériennes. Des colonies bien inférieures en taille à celles se développant sur boîte pourraient être détectées, ce qui autoriserait la diminution des temps d'incubation, l'augmentation des densités bactériennes sans recouvrement entre colonies, la visualisation des colonies moins favorisées par la croissance sur milieu nutritif *in vitro* et le criblage de toutes les bactéries se développant sur milieu sélectif en une seule fois.

Une méthode d'étalement sur lame a été testée et comparée avec celle sur boîte de Petri. Cent  $\mu\text{L}$  de la dilution  $10^{-6}$  d'une suspension d'électrotransformants sont mélangés directement sur lame à 900  $\mu\text{L}$  du milieu nutritif gélosé en surfusion sans pression de sélection. Parallèlement 100  $\mu\text{L}$  des dilutions  $10^{-4}$ ,  $10^{-5}$  et  $10^{-6}$  ont été étalés sur boîtes de Petri contenant du milieu agar sans pression de sélection. Le nombre d'UFC.mL<sup>-1</sup> ont été estimés après dénombrement des colonies. L'expérience a été reproduite 3 fois. Le nombre d'UFC.mL<sup>-1</sup> calculé est cohérent entre les deux méthodes d'étalement (**Tableau 13**). Un essai par ajout des cellules dans l'agar en surfusion avant étalement sur lame s'est révélé un peu moins concluant. Ce pré-mélange avant dépôt sur lame permettrait une meilleure homogénéisation des cellules dans le milieu nutritif. Une diminution du nombre d'UFC.mL<sup>-1</sup> de  $5.10^8$  à  $10^8$  a cependant été constatée. L'expérience n'a été reproduite qu'une seule fois, le mélange sur lame a néanmoins été préféré pour la suite des expérimentations.

La méthode de dénombrement des colonies sur lame ayant été validée, des premiers essais d'électrotransformation de pSM1885 ont été conduits avec une souche pure. L'objectif de ces expériences est d'une part de contrôler l'acquisition et la stabilisation du plasmide dans une souche environnementale, d'autre part de vérifier que l'expression du gène *gfp* est visualisable. Des espèces bactériennes présentes dans le sol sont naturellement fluorescentes, comme *Pseudomonas fluorescens*. Afin de vérifier que l'expression du gène *gfp* peut être détectable dans de telles souches, les essais d'électrotransformation de pSM1885 ont donc été réalisés avec la souche électrotransformable *P. fluorescens* N3 isolée par Cérémonie *et al.* (2004). *P. fluorescens* N3 a été rendue électrocompétente par lavages, soumise au choc électrique en présence ou absence du plasmide, puis étalée sur milieu sélectif en présence ou absence d'antibiotiques pour le dénombrement des transformants et des cellules réceptrices.





**Figure 13 : Intensité de fluorescence de *P. fluorescens* après acquisition de pSM1885**  
*Pseudomonas fluorescens* mis en présence (1) d'eau, (2) d'eau suivi d'un choc électrique, (3) de pSM1885 suivi d'un choc électrique, sont étalées sur TSA en absence (1, 2) ou en présence (3) des pressions de sélection adéquates, puis incubées 48 heures à 30°C. Les colonies sont observées sous microscope (A) sans filtre (B) avec filtre GFP

Seules les cellules de *Pseudomonas fluorescens* N3 ayant acquis le plasmide se développent sur milieu sélectif, le plasmide pSM1885 peut donc se répliquer dans la souche testée. La fréquence d'acquisition du plasmide *in vitro* est de  $1,5.10^{-4}$ , soit une fréquence 10 fois plus faible que celle décrite par Ceremonie *et al.* (2004). Suite à l'observation des lames sous la longueur d'onde d'excitation de la GFP, une différence d'intensité de fluorescence est constatée entre les bactéries soumises au choc électrique et les bactéries ayant acquis pSM1885 (**Figure 13**). Un gain de fluorescence est notamment observé pour les électrotransformants se développant sur milieu sélectif. Une surexpression de la fluorescence dans des souches naturellement fluorescentes est donc visuellement détectable.

Des essais d'électrotransformation de pSM1885 ont été conduits dans des cellules environnementales extraites sur gradient de Nycodenz. Les cellules ont été mises en présence de pSM1885, un choc électrique a été appliqué, puis les cellules ont été étalées sur lame par mélange avec du TSA en surfusion. De la DNAase a été ajoutée après électroporation pour éviter toute acquisition du plasmide par transformation naturelle, et ainsi ne sélectionner que les bactéries ayant acquis l'ADN par électrotransformation. Des témoins négatifs correspondant aux cellules environnementales seules (sans ajout d'ADN) soumises ou non à un choc électrique ont également été effectués pour vérifier la présence de bactéries naturellement résistantes aux antibiotiques. Après incubation des lames, des bactéries naturellement résistantes ont été détectées : des cellules environnementales sont capables de se développer en présence des deux antibiotiques tétracycline et streptomycine, bien qu'aucun plasmide n'ait été introduit. Ces colonies sont plus nombreuses si un choc électrique a eu lieu. Le choc électrique semble ainsi favoriser le développement d'une résistance aux deux antibiotiques chez les bactéries extraites du sol. Toutefois, le plus grand nombre de bactéries se développant sur milieu sélectif est observé pour les cellules ayant été mises en présence du plasmide avant choc électrique. Suite à l'observation des électrotransformants croissant sur milieu sélectif sous microscope équipé d'un filtre GFP, différentes intensités de fluorescence sont constatées suivant les colonies. Plusieurs questions se posent quant à ces différentes intensités de fluorescence : sont-elles dues à la fluorescence naturelle des bactéries ou à l'acquisition du plasmide ? Quelle sera l'intensité de fluorescence des colonies des bactéries ayant acquis le plasmide et non naturellement fluorescentes ? Qu'en est-il des bactéries exprimant faiblement la protéine GFP ? Un seuil d'intensité de fluorescence permettant la discrimination entre vrais électrotransformants et cellules naturellement fluorescentes est nécessaire. La définition de ce seuil nécessiterait des expérimentations complémentaires avec des souches pures électrotransformables non naturellement fluorescentes.

**Tableau 14 :** Efficacité de la cassette toxique de pFEB98 chez différentes espèces bactériennes

Espèces bactériennes	Concentrations en IPTG ( $\mu\text{g.mL}^{-1}$ )		
	250	500	1000
<i>E. coli</i> DH5 $\alpha$ (pFEB98)	$5.10^{-7}$	$5.10^{-7}$	$9.10^{-7}0$
<i>A. baylyi</i> BD413 (pFEB98)	$10^{-5}$	$5.10^{-6}$	ND <sup>a</sup>
<i>P. fluorescens</i> N3 (pFEB98)	$10^0$	$10^0$	$10^0$

L'efficacité de pFEB98 chez différentes espèces bactériennes est estimée par le calcul d'un taux d'échappement après induction par l'IPTG.

<sup>a</sup> ND : non déterminé

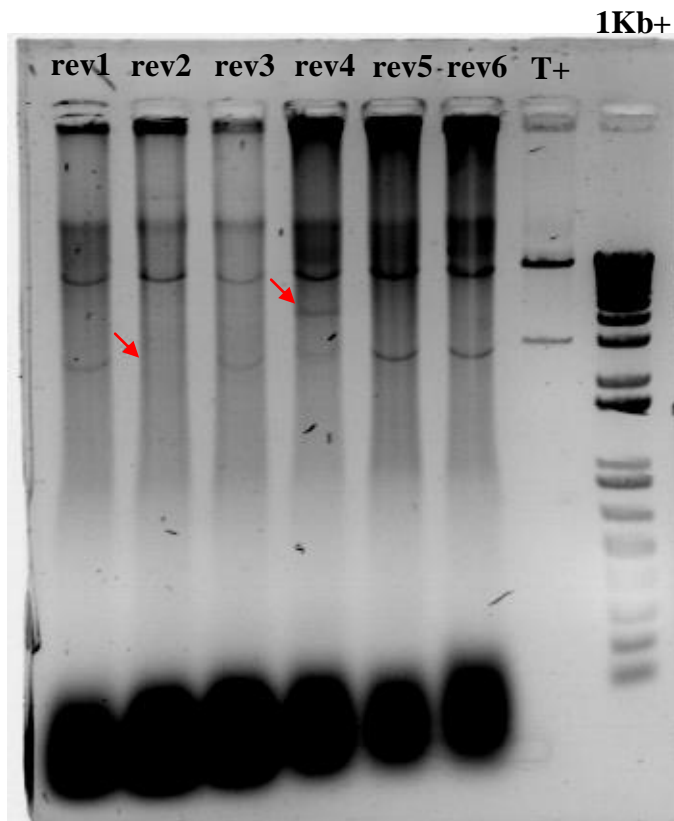
### 3.2 Efficacité de gènes toxiques dans différentes souches environnementales

Plusieurs souches environnementales sont capables d'acquérir l'ADN extracellulaire à des fréquences élevées. Leur utilisation comme souche réceptrice implique néanmoins qu'elles soient sensibles à l'expression toxique inductible du système de contre sélection. La cassette toxique inductible actuellement construite est constituée des gènes *relF* et *colE3*. L'activité toxique de *colE3* a été observée par Diaz *et al.* (1994) chez de nombreuses souches environnementales. Le gène *relF*, quant à lui est efficace chez *E. coli* (Knudsen *et al.*, 1995; Knudsen & Karlstrom, 1991). Son mode d'action laisse envisager un spectre d'action plus large. Son efficacité toxique n'a cependant pas encore été vérifiée dans ces souches environnementales.

Nous avons voulu tester l'action toxique de *relF* dans les deux souches *Pseudomonas fluorescens* et *Acinetobacter baylyi*. La première est capable d'internaliser l'ADN par électrotransformation dans le sol à une fréquence de  $10^{-4}$ . La seconde peut également acquérir l'ADN extracellulaire par transformation naturelle à des fréquences importantes. Ces deux souches présentant des voies d'acquisition d'ADN différentes pourraient être toutes deux potentiellement intéressantes comme souche réceptrice. Le plasmide à large spectre d'hôte pFEB98 (contenant deux gènes *relF*) a été transformé dans chacune de ces deux souches ainsi que dans une souche *E. coli* utilisée comme témoin positif, puis son action toxique dans la cellule hôte vérifiée par un test d'efficacité.

L'induction de l'expression de *relF* n'a aucun effet toxique chez *P. fluorescens*, le taux d'échappement étant de  $10^0$  (**Tableau 14**). Le promoteur régulant l'expression de *relF* a déjà été utilisé avec succès chez *Pseudomonas* (Jensen *et al.*, 1993), cette absence d'effet létal ne peut donc être expliqué par un problème de fonctionnement du système régulateur dans la cellule hôte. Par ailleurs, le plasmide isolé de *P. fluorescens* et retransformé dans *E. coli* DH5 $\alpha$  entraîne une mort cellulaire de *E. coli* lorsqu'il est induit. Un système toxique fonctionnel a donc bien été introduit dans *P. fluorescens*, la protéine RelF ne semble cependant avoir aucun effet létal sur la cellule hôte.

L'induction de l'expression de *relF* chez *A. baylyi* entraîne par contre la mort de la cellule hôte et ce avec un taux d'échappement atteignant  $10^{-5}$ . Ce taux d'échappement est 10 fois supérieur à celui observé chez *E. coli*. Pour comprendre la raison de l'échappement de certaines cellules d'*A. baylyi* contenant pFEB98 à l'expression toxique de *relF*, six bactéries échappant à l'induction de l'expression toxique (nommés ici des révertants) ont été



**Figure 14 :** Profil de restriction de plasmides extraits de *A. Baylyi* (pFEB98) échappant à l'induction de l'expression toxique du gène *relF*

Les plasmides de 6 clones (**rev1 à rev6**) de *A. baylyi* échappant à l'induction de l'expression de *relF* (révertants) ont été extraits, digérés par HindIII et migrés sur gel d'agarose en présence de pFEB98 extrait de *E. coli* DH5α (**T+**) ayant subi le même traitement enzymatique et du marqueur de taille **1Kb+**. Les variations du profil de digestion des plasmides extraits des révertants comparé à celui du pFEB98 sont marquées par une flèche rouge.

sélectionnés de manière aléatoire pour la suite des tests. Tous les 6 possèdent un plasmide noté Rev1 à Rev6, dont le profil de restriction est variable (**Figure 14**).

Après digestion des 6 plasmides par différentes enzymes de restriction (dont HindIII) et comparaison des profils de restriction obtenus avec celui du témoin pFEB98, deux événements ont pu être identifiés : une délétion et une insertion de séquence d'ADN sur respectivement Rev2 et Rev4. Ainsi, certaines cellules d'*A. baylyi* (pFeb98) échapperaient à l'induction du système suicide par des événements mutationnels d'insertion ou délétion de séquence d'ADN au niveau de la construction suicide. Ces événements sont en cohérence avec la carte de pFEB98, sur laquelle deux séquences homologues, les gènes *bla* (gène de résistance à l'ampicilline), sont relevés. Ces deux zones homologues peuvent en effet être le lieu de recombinaison puis de délétion de la séquence d'ADN intermédiaire, ce qui entraînerait l'élimination d'un gène suicide sur les deux initialement présents. L'existence de ce phénomène de recombinaison chez *A. baylyi* est également confortée par l'aptitude de la souche à recombiner l'ADN à des fréquences élevées (Nielsen *et al.*, 1997b).

#### **4 CONCLUSION**

Dans notre approche de capture de gènes, certaines propriétés sont requises pour l'exploitation d'une espèce bactérienne comme souche réceptrice, dont la sensibilité de la bactérie à l'expression de la cassette toxique, la capacité à acquérir l'ADN de la microflore tellurique et à recombiner cet ADN qui a pénétré dans la cellule.

Les travaux présentés dans ce chapitre mettent en valeur l'importance du contrôle de la sensibilité de la souche bactérienne à l'expression toxique, même si le mode d'action de la protéine exprimée par le gène toxique laisse envisager un spectre d'action large. Ainsi le gène *relF*, codant une porine, n'entraîne aucune action létale sur *Pseudomonas fluorescens* N3. Deux stratégies peuvent alors être adoptées : soit le choix de l'espèce bactérienne comme souche réceptrice est conditionnée par sa sensibilité à l'expression de la cassette toxique construite, soit la cassette toxique est définie en fonction de la sensibilité de l'espèce bactérienne choisie aux gènes toxiques testés.

Par ailleurs, l'activité recombinogène de la souche réceptrice peut influencer l'efficacité de la construction toxique utilisée si des zones d'homologies sont présentes sur la construction. Ainsi le plasmide pFEB98 est sujet à des événements mutationnels d'insertion ou délétion dans *Acinetobacter baylyi*. L'emploi de souches hautement recombinogènes implique donc une grande prudence lors de la définition de la cassette toxique qui fera partie

intégrante du système de contre sélection positive, la moindre homologie pouvant affecter l'efficacité globale de l'outil de capture de gène.

*Acinetobacter baylyi*, souche présente dans l'environnement sol, naturellement transformable, et recombino-gène, est également sensible à l'induction de l'expression du gène *relF*. Cette souche est donc très intéressante pour son exploitation comme souche réceptrice. Quelques modifications ultérieures, notamment par l'ajout d'un plasmide comportant une cassette toxique encadrée par les régions ADN cibles, permettra son utilisation *in situ* pour la capture des informations génétiques ciblées de la microflore tellurique, sous condition de compétence.

L'électrotransformation est une voie d'acquisition intéressante de l'ADN, puisqu'elle ne requière pas de stade de compétence. A ce jour seulement deux souches électrotransformables ont été isolées du sol, dont *Pseudomonas fluorescens* N3 capable d'acquérir de l'ADN à haute fréquence. Nos travaux ont montré l'inefficacité de *relF* dans cette souche. Le développement d'une méthode pour isoler des bactéries hautement électrotransformable a été initiée, afin d'accroître la collection d'espèces bactériennes telluriques électrotransformables. Des travaux complémentaires sont cependant nécessaires pour définir le seuil de détection des cellules ayant acquis l'ADN par choc électrique.

## CONCLUSIONS ET PERSPECTIVES

La capture *in situ* de l'ADN de la microflore tellurique est un concept novateur pour accéder à la diversité génétique des gènes (ou opérons) ciblés des bactéries présentes dans l'environnement étudié. Un des intérêts de cette approche est le contournement des étapes d'extraction et purification de l'ADN métagénomique de l'échantillon étudié, étapes à l'origine d'une perte d'informations génétiques. La mise en pratique de ce concept a néanmoins nécessité de nombreux développements technologiques, ainsi que de nombreuses précautions pour rendre l'outil de piégeage opérationnel.

Un système de contre sélection efficace a notamment été développé pour capturer même les gènes peu représentés dans le métagénome. Des premiers travaux expérimentaux effectués pour la capture de la diversité des gènes codant l'enzyme phenol hydroxylase ont en effet montré que les événements de transformation / recombinaison de l'ADN du métagénome dans la souche réceptrice modifiée pouvaient avoir lieu à des fréquences inférieures à  $10^{-7}$ . La détection de ces événements implique donc l'emploi d'une cassette toxique dont le taux d'échappement soit inférieur à  $10^{-7}$ . De telles efficacités peuvent être atteintes par la construction d'une cassette constituée de plusieurs gènes toxiques. L'emploi de deux gènes toxiques différents nous a ainsi permis d'atteindre un taux d'échappement minimum de  $10^{-7}$ .

La construction d'une telle cassette toxique a cependant tout d'abord nécessité le choix d'un système régulateur inductible adapté au gène toxique utilisé. Nos travaux ont en effet mis en valeur l'importance du choix du système régulateur en fonction de la toxicité exprimée par le gène utilisé. Ainsi le gène *relF* pour être efficace doit être fortement exprimé en condition d'induction, tout en ayant une expression résiduelle en condition non inductrice suffisamment faible pour éviter une réponse hormétique de la cellule hôte. Le gène *colE3*, dont la production d'une faible quantité de protéines suffit à conduire à la mort de la souche hôte, a nécessité le développement d'un nouveau système de régulation inductible, inhibant toute expression résiduelle en absence d'inducteur.



La stratégie de capture étant basée sur la recombinaison par homologie de séquence entre régions cibles clonées et ADN du métagénome, un soin particulier doit également être apporté à la construction pour éviter tout événement de recombinaison autre que ceux attendus. Ainsi la présence de séquences homologues sur le plasmide comportant la cassette toxique peut conduire à la délétion de la cassette, tout comme la présence d'homologies entre les deux régions cibles, affectant ainsi l'efficacité du système de contre sélection.

Le système développé est actuellement exploitable dans *E. coli*. Cette espèce bactérienne présente l'avantage d'être facilement manipulable génétiquement et autorise l'emploi de systèmes lambda red gam pour activer la recombinaison. Bien que la souche *E. coli* utilisée, issue de nombreuses modifications génétiques, soit moins compétitive par rapport aux espèces autochtones lors de son inoculation dans l'environnement sol (Recorbet *et al.*, 1993), son emploi *in situ* pour capturer l'ADN de la microflore tellurique est tout à fait envisageable via l'électrotransformation. Son introduction dans le sol suivi d'un choc électrique permettra l'entrée de l'ADN dans la cellule hôte, sans étape de colonisation, ni de développement d'une compétence *in situ*. L'électrotransformation permettra ainsi un gain de temps considérable pour l'obtention des résultats de capture. Dans cette stratégie, l'ADN de la microflore tellurique doit cependant être au préalable libéré pour être disponible à la capture. Les traitements lytiques *in situ* se sont révélés très efficaces, avec plus de 90 % des bactéries lysées par la combinaison d'un traitement chimique avec le traitement par agitation de billes d'acier (Howeler *et al.*, 2003). Cette méthode présente l'avantage de déstructurer le sol conjointement à la lyse cellulaire, offrant ainsi une possibilité d'accès à l'ensemble de la diversité bactérienne. L'ADN libéré est alors protégé des activités nucléasiques grâce au lysat cellulaire (Ceremonie *et al.*, 2006; Nielsen *et al.*, 2000), et reste disponible pour l'électrotransformation (Ceremonie *et al.*, 2006). Comme nous l'avons cependant auparavant précisé, cette technique est très destructrice, générant des fragments d'ADN entre 0,5 et 20 kb (Cullen & Hirsch, 1998; Frostegard *et al.*, 1999; Ogram *et al.*, 1987).

L'emploi d'une méthode lytique tout aussi efficace mais moins destructive telle que la lyse chimique / enzymatique ne permet pas la déstructuration du sol. L'accès à la diversité des informations génétiques contenues dans les différents compartiments du sol implique donc l'utilisation d'une espèce bactérienne capable de coloniser l'environnement sol. Des espèces naturellement transformables ou électrotransformables directement isolées de l'environnement sol seraient donc particulièrement intéressantes. Ces espèces doivent cependant présenter certaines propriétés pour être exploitées comme souches réceptrices, dont

la sensibilité à la cassette toxique définie, une haute (électro) transformabilité *in situ*, ainsi qu'une activité recombinogène.

L'électrotransformation présente de nombreux avantages. Outre la capacité internaliser l'ADN sans besoin de développement de compétence, l'électroporation contribuerait à diminuer l'activité exonucléasique de RecD et ainsi favoriser la recombinaison. Peu d'espèces bactériennes électrotransformables ont à ce jour été isolées du sol. Dans ces travaux a été initiée une méthode pour détecter de nouvelles souches électrotransformables, des expérimentations complémentaires sont cependant nécessaires pour son application.

Par contre, de nombreuses espèces bactériennes présentes dans le sol ont été décrites comme naturellement transformables. Nos travaux ont montré qu'*Acinetobacter baylyi* serait une souche réceptrice potentielle (sous condition de sa sensibilité à l'expression de *colE3*). Bien que la souche perde rapidement sa compétence *in situ*, celle-ci peut être réactivée par l'ajout de nutriments *in situ* (Nielsen *et al.*, 1997a). Cet ajout de nutriments pourrait également favoriser la colonisation dans les porosités du sol plus petites de la souche (Dechesne *et al.*, 2005). Le mode de colonisation est cependant variable d'une espèce à l'autre (Dechesne *et al.*, 2003; Grundmann *et al.*, 2001; Pallud *et al.*, 2004). La multiplicité d'utilisation de souches réceptrices capables de coloniser différents compartiments du sol étendrait l'accès à une diversité plus complète.

Dans ces travaux, nous nous étions plus particulièrement intéressés à ces deux voies, transformation et électrotransformation, d'acquisition d'ADN. Il serait néanmoins intéressant d'étendre le système aux voies d'acquisition par conjugaison ou transduction. Les phages par exemple sont abondants dans l'environnement. Le dénombrement direct des phages avec un microscope à transmission électronique a permis d'évaluer la densité phagique d'un sol à environ  $1,5 \times 10^8 \cdot g^{-1}$  (Ashelford *et al.*, 2003). La capture de l'information génétique via un vecteur phagique requière un phage tempéré à large spectre d'hôte et à transduction généralisée. De tels phages ont déjà été isolés (Jensen *et al.*, 1998). Par ailleurs des événements de transduction ont été mis en évidence dans le sol (Zeph *et al.*, 1988). Des expérimentations seraient nécessaires pour valider la pertinence de leur emploi dans notre approche. Dans le cas de la conjugaison, deux stratégies pourraient être optées : soit la souche réceptrice possède le matériel nécessaire à la conjugaison, un rétro transfert permettrait alors la capture de l'ADN de la microflore tellurique (Szpirer *et al.*, 1999). Soit l'acquisition de l'ADN de la microflore tellurique sera dépendante des bactéries indigènes présentant le matériel nécessaire à la conjugaison, ce qui permettrait d'en apprendre un peu plus sur la prévalence des plasmides conjuguatifs dans la microflore tellurique.



## CONCLUSION GENERALE

Alors que seule une infime proportion de la diversité bactérienne est accessible par les approches culturales, les approches métagénomiques développées ces dernières années ont élargi les données sur les bactéries non cultivées, permettant de nombreuses avancées scientifiques tant au niveau fondamental qu'appliqué. Ces approches, basées sur l'extraction préalable de l'ensemble des génomes (métagénome) des bactéries présentes dans l'échantillon étudié, ne permettent cependant l'accès qu'à une fraction des informations génétiques contenues dans l'environnement sol. Près de 30 à 50 % des informations génétiques sont perdues lors des processus d'extraction et purification de l'ADN des composants du sol, sans compter celles perdues lors des voies d'analyses ultérieures de l'ADN. Ainsi, la construction des banques métagénomiques peut s'accompagner de la perte d'ADN due à une purification insuffisante de l'ADN, ou à des limitations techniques liées à la taille des banques requise pour cloner la totalité de l'information génétique extraite.

Aux cours des différents travaux présentés dans cette thèse, nous avons cherché les moyens d'accéder à une part plus large de la diversité génétique. Une première stratégie visait à améliorer la qualité de l'ADN extrait. L'augmentation de la taille des ADN extraits permettrait notamment la diminution du nombre de clones requis lors des opérations de clonage et améliorerait ainsi la couverture des informations génétiques présentes dans l'échantillon de sol étudié. Ces considérations nous ont ainsi conduits à étudier les facteurs pouvant influencer la qualité de l'ADN extrait. Les résultats de ces travaux ont montré que les méthodes d'extraction n'expliquent pas à elles seules la parfois faible taille des ADN extraits, mais un autre facteur lié à la physiologie cellulaire aurait un impact non négligeable. Dans l'hypothèse alors proposée, les bactéries en phase de dormance n'alloueraient pas d'énergie dans la conservation de l'intégrité de leur ADN génomique. Bien que la stimulation métabolique apportée dans un sol pauvre en matière organique ait permis d'augmenter la taille des ADN extraits, elle s'accompagne d'une modification de la diversité génétique. L'amélioration de la qualité de l'ADN semblait ainsi indissociable d'un biais vers l'enrichissement d'une partie de la population bactérienne.

Bien que les enrichissements s'avèrent particulièrement intéressants pour l'étude de populations ciblées, une autre stratégie pour accéder aux données manquantes sur la diversité bactérienne a été approfondie. Cette deuxième stratégie visait à contourner les biais liés à l'extraction en capturant directement *in situ* l'ADN de la microflore tellurique. Cette stratégie reposant sur la conception d'une souche réceptrice capable de piéger l'ADN de la microflore, a nécessité de nombreux développements technologiques. Cette souche devait en effet être capable d'acquérir et d'intégrer l'ADN de la microflore tellurique par recombinaison dans son génome afin de le stabiliser. Les souches réceptrices ayant acquis cet ADN devaient pouvoir être sélectionnées spécifiquement. Au cours des développements technologiques, ce dernier point s'est révélé particulièrement délicat : les événements de capture pouvant avoir lieu à des fréquences très faibles, un système de sélection capable de détecter ces événements était indispensable.

Un système de contre sélection positive a ainsi été développé pour tuer toutes souches réceptrices n'ayant pas piégé d'ADN. Le système développé est constitué de deux sites de recombinaison définis dans les régions conservées des gènes dont la diversité de séquences génétiques est recherchée, encadrant une cassette toxique inductible. Suite à une similarité de séquences entre l'ADN de la microflore tellurique et les sites de recombinaison, une double recombinaison conduit au remplacement de la cassette toxique inductible par l'ADN de la microflore tellurique. Seules les souches ayant effectué ce remplacement survivent en présence des inducteurs de l'expression des gènes toxiques de la cassette. Ce système simple en théorie a cependant nécessité de nombreux développements et de nombreuses précautions, premièrement pour manipuler les gènes toxiques dans un système d'expression inductible, deuxièmement pour réaliser une cassette toxique efficace (avec un faible taux d'échappement), et troisièmement pour la manipulation de souches recombinogènes. Les moindres similarités de séquences non contrôlées peuvent en effet conduire à l'altération de l'efficacité du système de contre sélection.

Des premiers essais expérimentaux *in vitro* de capture d'ADN conduits avec *E. coli* comme souche réceptrice ont mis en valeur l'importance de l'efficacité de la cassette toxique. Ces essais ont mené à la construction d'une cassette toxique constituée de deux gènes différents chacun régulé de manière indépendante. Le taux d'échappement actuel de la cassette dans les conditions testées est de  $10^{-7}$ , pouvant potentiellement encore atteindre des taux d'échappement plus faibles après quelques optimisations génétiques. Cette cassette peut être clonée dans d'autres vecteurs, chacun comportant des sites de recombinaison définis en

fonction de la diversité des gènes recherchés. Les premières études effectuées avec les gènes codant les enzymes phénols hydroxylases peuvent donc s'étendre à l'étude de la diversité d'autres gènes, sous condition de présence de régions suffisamment conservées en amont et aval du gène ciblé pour la définition des sites de recombinaison.

Les travaux initiés nécessitent des tests complémentaires *in vitro* sur la limite de divergence entre séquences acceptées pour l'intégration de l'ADN cible dans le système de contre sélection. L'étude de ce paramètre déterminera le degré de diversité (richesse) des séquences nucléotidiques qui pourra être capturé par l'outil développé. Par ailleurs, des études réalisées sur la quantité de molécules d'ADN cibles minimales présentes dans le métagénome pouvant être capturées par la souche réceptrice, donnera une idée sur l'abondance nécessaire des gènes recherchés pour leur capture. Enfin, l'introduction de l'outil de capture *in situ* permettra la mise en application du concept proposé dans cette thèse, à savoir l'accès à l'information génétique de la microflore tellurique sans étape d'extraction de l'ADN des bactéries de la matrice sol, étape qui occasionne la perte de l'information génétique.

La multiplicité des approches permet de gagner en informations sur la diversité bactérienne tellurique présente dans un environnement. Les approches utilisées pour analyser la diversité bactérienne présente dans un sol, à ma connaissance, font toutes intervenir une étape préalable d'extraction de l'ADN de la matrice sol. Même si l'outil de capture proposé présente des biais liés à la diminution de la fréquence de transformation *in situ*, ainsi que des biais liés à la recombinaison de séquences similaires à celles clonées dans la souche réceptrice, il présente la potentialité d'accéder à un nouveau degré de diversité en contournant cette étape d'extraction. Ceci pourrait être confirmé par la comparaison des informations récoltées par l'outil de capture et celles contenues dans les banques métagénomiques.

Par ailleurs, l'outil développé présente en lui-même une multiplicité d'approches via l'emploi de différentes espèces bactériennes aptes à capturer l'ADN, de différentes voies d'acquisition d'ADN et de stratégies différentes pour capturer cet ADN. La capture de l'information génétique dans un sol soumis à un traitement physique, chimique et enzymatique pour déstructurer le sol conjointement à la lyse cellulaire permettra l'accès à une information plus importante. Une autre stratégie visant à conserver la structure du sol permettrait d'accéder aux informations génétiques présente dans les différentes structures du sol (macroagrégats / microagrégats) suivant le mode de colonisation de la souche réceptrice introduite et des artifices employés pour favoriser sa colonisation dans certaines fractions du sol (ajout de nutriments).

Outre le contexte d'étude pour lequel le système a été développé, de nombreuses autres applications de l'outil peuvent être envisagées, tels que dans les contextes d'endiguement génétique où de très faibles taux d'échappement sont recherchés, ou dans le contexte d'étude des transferts d'ADN entre organismes génétiquement modifiés et bactéries afin de savoir si de tels transferts sont possibles dans l'environnement. Le faible taux d'échappement de la cassette toxique développée, la haute fréquence de transformation de *E. coli* par électrotransformation, sa forte activité recombinogène par l'emploi de systèmes red gam et l'introduction de sites de recombinaison homologues au transgène par exemple suivi en font un outil très intéressant pour détecter de tels événements dans l'environnement.





## REFERENCES BIBLIOGRAPHIQUES

**Aakra, A., Hesselsoe, M. & Bakken, L. R. (2000).** Surface Attachment of Ammonia-Oxidizing Bacteria in Soil. *Microb Ecol* **39**, 222-235.

**Afif, H., Allali, N., Couturier, M. & Van Melderren, L. (2001).** The ratio between CcdA and CcdB modulates the transcriptional repression of the ccd poison-antidote system. *Mol Microbiol* **41**, 73-82.

**Ahrenholtz, I., Lorenz, M. G. & Wackernagel, W. (1994).** A conditional suicide system in *Escherichia coli* based on the intracellular degradation of DNA. *Appl Environ Microbiol* **60**, 3746-3751.

**Aizenman, E., Engelberg-Kulka, H. & Glaser, G. (1996).** An *Escherichia coli* chromosomal "addiction module" regulated by guanosine [corrected] 3',5'-bispyrophosphate: a model for programmed bacterial cell death. *Proc Natl Acad Sci U S A* **93**, 6059-6063.

**Alexander, M. (1977).** *Introduction to soil microbiology*, second edn. United States of America: John Wiley & Sons, Inc.

**Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990).** Basic local alignment search tool. *J Mol Biol* **215**, 403-410.

**Amann, R. I., Ludwig, W. & Schleifer, K. H. (1995).** Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* **59**, 143-169.

**Amundsen, S. K., Taylor, A. F., Chaudhury, A. M. & Smith, G. R. (1986).** recD: the gene for an essential third subunit of exonuclease V. *Proc Natl Acad Sci U S A* **83**, 5558-5562.

**Andersen, J. B., Sternberg, C., Poulsen, L. K., Bjorn, S. P., Givskov, M. & Molin, S. (1998).** New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl Environ Microbiol* **64**, 2240-2246.

**Anthony, L. C., Suzuki, H. & Filutowicz, M. (2004).** Tightly regulated vectors for the cloning and expression of toxic genes. *J Microbiol Methods* **58**, 243-250.

**Ashelford, K. E., Day, M. J. & Fry, J. C. (2003).** Elevated abundance of bacteriophage infecting bacteria in soil. *Appl Environ Microbiol* **69**, 285-289.

**Bakken, L. R. (1985).** Separation and Purification of Bacteria from Soil. *Appl Environ Microbiol* **49**, 1482-1487.

**Barra Caracciolo, A., Grenni, P., Cupo, C. & Rossetti, S. (2005).** In situ analysis of native microbial communities in complex samples with high particulate loads. *FEMS Microbiol Lett* **253**, 55-58.

- Bej, A. K., Perlin, M. H. & Atlas, R. M. (1988).** Model suicide vector for containment of genetically engineered microorganisms. *Appl Environ Microbiol* **54**, 2472-2477.
- Beja, O., Suzuki, M. T., Koonin, E. V. & other authors (2000).** Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ Microbiol* **2**, 516-529.
- Bergsma-Vlami, M., Prins, M. E., Staats, M. & Raaijmakers, J. M. (2005).** Assessment of genotypic diversity of antibiotic-producing pseudomonas species in the rhizosphere by denaturing gradient gel electrophoresis. *Appl Environ Microbiol* **71**, 993-1003.
- Berry, A. E., Chiocchini, C., Selby, T., Sosio, M. & Wellington, E. M. (2003).** Isolation of high molecular weight DNA from soil for cloning into BAC vectors. *FEMS Microbiol Lett* **223**, 15-20.
- Bertolla, F. & Simonet, P. (1999).** Horizontal gene transfers in the environment: natural transformation as a putative process for gene transfers between transgenic plants and microorganisms. *Res Microbiol* **150**, 375-384.
- Bertolla, F., Pepin, R., Passelegue-Robe, E., Paget, E., Simkin, A., Nesme, X. & Simonet, P. (2000).** Plant genome complexity may be a factor limiting in situ the transfer of transgenic plant genes to the phytopathogen *Ralstonia solanacearum*. *Appl Environ Microbiol* **66**, 4161-4167.
- Bertrand, H., Poly, F., Van, V. T., Lombard, N., Nalin, R., Vogel, T. M. & Simonet, P. (2005).** High molecular weight DNA recovery from soils prerequisite for biotechnological metagenomic library construction. *J Microbiol Methods* **62**, 1-11.
- Best, M. (1995).** Classification of rocks and their abundances on the earth. In *Rocks physics and phase relation: a handbook of physical contents* pp. 1-7. Edited by T. Ahrens. United States of America: the American Geophysical Union.
- Bockelmann, U., Szewzyk, U. & Grohmann, E. (2003).** A new enzymatic method for the detachment of particle associated soil bacteria. *J Microbiol Methods* **55**, 201-211.
- Bowers, L. M., Lapoint, K., Anthony, L., Pluciennik, A. & Filutowicz, M. (2004).** Bacterial expression system with tightly regulated gene expression and plasmid copy number. *Gene* **340**, 11-18.
- Brady, S. F., Chao, C. J. & Clardy, J. (2002).** New natural product families from an environmental DNA (eDNA) gene cluster. *J Am Chem Soc* **124**, 9968-9969.
- Breitbart, M., Hewson, I., Felts, B., Mahaffy, J. M., Nulton, J., Salamon, P. & Rohwer, F. (2003).** Metagenomic analyses of an uncultured viral community from human feces. *J Bacteriol* **185**, 6220-6223.
- Bruneau, P. M., Davidson, D. A., Grieve, I. C., Young, I. M. & Nunan, N. (2005).** The effects of soil horizons and faunal excrement on bacterial distribution in an upland grassland soil. *FEMS Microbiol Ecol* **52**, 139-144.

- Calabrese, E. J. (2004).** Hormesis: a revolution in toxicology, risk assessment and medicine. *EMBO Rep* **5** Spec No, S37-40.
- Ceremonie, H., Buret, F., Simonet, P. & Vogel, T. M. (2004).** Isolation of lightning-competent soil bacteria. *Appl Environ Microbiol* **70**, 6342-6346.
- Ceremonie, H., Buret, F., Simonet, P. & Vogel, T. M. (2006).** Natural *Pseudomonas* sp. strain N3 in artificial soil microcosms. *Appl Environ Microbiol* **72**, 2385-2389.
- Chamier, B., Lorenz, M. G. & Wackernagel, W. (1993).** Natural Transformation of *Acinetobacter calcoaceticus* by Plasmid DNA Adsorbed on Sand and Groundwater Aquifer Material. *Appl Environ Microbiol* **59**, 1662-1667.
- Chang, A. C. & Cohen, S. N. (1978).** Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J Bacteriol* **134**, 1141-1156.
- Chaudhury, A. M. & Smith, G. R. (1984).** A new class of *Escherichia coli* recBC mutants: implications for the role of RecBC enzyme in homologous recombination. *Proc Natl Acad Sci U S A* **81**, 7850-7854.
- Chaveroche, M. K., Ghigo, J. M. & d'Enfert, C. (2000).** A rapid method for efficient gene replacement in the filamentous fungus *Aspergillus nidulans*. *Nucleic Acids Res* **28**, E97.
- Chen, D. Q., Huang, S. S. & Lu, Y. J. (2006).** Efficient transformation of *Legionella pneumophila* by high-voltage electroporation. *Microbiol Res* **161**, 246-251.
- Chen, G. & Zhu, H. (2005).** Bacterial adhesion to silica sand as related to Gibbs energy variations. *Colloids Surf B Biointerfaces* **44**, 41-48.
- Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. & Thompson, J. D. (2003).** Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* **31**, 3497-3500.
- Chenu, C. (1993).** Clay- or sand-polysaccharide associations as models for the interface between micro-organisms and soil: water related properties and microstructure. *Geoderma* **56**, 143-156.
- Christensen, H., Hansen, M. & Sorensen, J. (1999).** Counting and size classification of active soil bacteria by fluorescence in situ hybridization with an rRNA oligonucleotide probe. *Appl Environ Microbiol* **65**, 1753-1761.
- Copeland, N. G., Jenkins, N. A. & Court, D. L. (2001).** Recombineering: a powerful new tool for mouse functional genomics. *Nat Rev Genet* **2**, 769-779.
- Courtois, S., Frostegard, A., Goransson, P., Depret, G., Jeannin, P. & Simonet, P. (2001).** Quantification of bacterial subgroups in soil: comparison of DNA extracted directly from soil or from cells previously released by density gradient centrifugation. *Environ Microbiol* **3**, 431-439.

- Courtois, S., Cappellano, C. M., Ball, M. & other authors (2003).** Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl Environ Microbiol* **69**, 49-55.
- Couturier, M., Bahassi el, M. & Van Melderren, L. (1998).** Bacterial death by DNA gyrase poisoning. *Trends Microbiol* **6**, 269-275.
- Cullen, D. W. & Hirsch, P. R. (1998).** Simple and rapid method for direct extraction of microbial DNA from soil for PCR. *Soil Biology and Biochemistry* **30**, 983-993.
- Curtis, T. P. & Sloan, W. T. (2004).** Prokaryotic diversity and its limits: microbial community structure in nature and implications for microbial ecology. *Curr Opin Microbiol* **7**, 221-226.
- Dabert, P. & Smith, G. R. (1997).** Gene replacement with linear DNA fragments in wild-type *Escherichia coli*: enhancement by Chi sites. *Genetics* **145**, 877-889.
- Daly, M. J., Ouyang, L., Fuchs, P. & Minton, K. W. (1994).** In vivo damage and recA-dependent repair of plasmid and chromosomal DNA in the radiation-resistant bacterium *Deinococcus radiodurans*. *J Bacteriol* **176**, 3508-3517.
- Daniel, R. (2004).** The soil metagenome--a rich resource for the discovery of novel natural products. *Curr Opin Biotechnol* **15**, 199-204.
- Daniel, R. (2005).** The metagenomics of soil. *Nat Rev Microbiol* **3**, 470-478.
- Datsenko, K. A. & Wanner, B. L. (2000).** One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640-6645.
- Datta, S., Costantino, N. & Court, D. L. (2006).** A set of recombineering plasmids for gram-negative bacteria. *Gene* **379**, 109-115.
- Davison, J. (1999).** Genetic exchange between bacteria in the environment. *Plasmid* **42**, 73-91.
- De Fede, K. L. & Sexstone, A. J. (2001).** Differential response of size-fractionated soil bacteria in BIOLOG(R) microtitre plates. *Soil Biology and Biochemistry* **33**, 1547-1554.
- de Vries, J., Meier, P. & Wackernagel, W. (2001).** The natural transformation of the soil bacteria *Pseudomonas stutzeri* and *Acinetobacter* sp. by transgenic plant DNA strictly depends on homologous sequences in the recipient cells. *FEMS Microbiol Lett* **195**, 211-215.
- de Vries, J. & Wackernagel, W. (2002).** Integration of foreign DNA during natural transformation of *Acinetobacter* sp. by homology-facilitated illegitimate recombination. *Proc Natl Acad Sci U S A* **99**, 2094-2099.
- Dechesne, A., Pallud, C., Debouzie, D., Flandrois, J. P., Vogel, T. M., Gaudet, J. P. & Grundmann, G. L. (2003).** A novel method for characterizing the microscale 3D spatial distribution of bacteria in soil. *Soil Biology and Biochemistry* **35**, 1537-1546.

- Dechesne, A., Pallud, C., Bertolla, F. & Grundmann, G. L. (2005).** Impact of the microscale distribution of a *Pseudomonas* strain introduced into soil on potential contacts with indigenous bacteria. *Appl Environ Microbiol* **71**, 8123-8131.
- Delmas, S. & Matic, I. (2005).** Cellular response to horizontally transferred DNA in *Escherichia coli* is tuned by DNA repair systems. *DNA Repair (Amst)* **4**, 221-229.
- Demain, A. L. (1999).** Pharmaceutically active secondary metabolites of microorganisms. *Appl Microbiol Biotechnol* **52**, 455-463.
- Demaneche, S., Bertolla, F., Buret, F., Nalin, R., Sailland, A., Auriol, P., Vogel, T. M. & Simonet, P. (2001a).** Laboratory-scale evidence for lightning-mediated gene transfer in soil. *Appl Environ Microbiol* **67**, 3440-3444.
- Demaneche, S., Jocteur-Monrozier, L., Quiquampoix, H. & Simonet, P. (2001b).** Evaluation of biological and physical protection against nuclease degradation of clay-bound plasmid DNA. *Appl Environ Microbiol* **67**, 293-299.
- Demaneche, S., Kay, E., Gourbiere, F. & Simonet, P. (2001c).** Natural transformation of *Pseudomonas fluorescens* and *Agrobacterium tumefaciens* in soil. *Appl Environ Microbiol* **67**, 2617-2621.
- Dercourt, J. & Paquet, J. (1981).** *Géologie: objets et méthodes*, 5 edn. Paris: Bordas.
- Diaz, E., Munthali, M., de Lorenzo, V. & Timmis, K. N. (1994).** Universal barrier to lateral spread of specific genes among microorganisms. *Mol Microbiol* **13**, 855-861.
- Dong, D., Yan, A., Liu, H., Zhang, X. & Xu, Y. (2006).** Removal of humic substances from soil DNA using aluminium sulfate. *J Microbiol Methods* **66**, 217-222.
- Donlan, R. M. (2002).** Biofilms: microbial life on surfaces. *Emerg Infect Dis* **8**, 881-890.
- Doyle, R. J. (2000).** Contribution of the hydrophobic effect to microbial infection. *Microbes Infect* **2**, 391-400.
- Draghi, J. A. & Turner, P. E. (2006).** DNA secretion and gene-level selection in bacteria. *Microbiology* **152**, 2683-2688.
- Duchaufour, P. (1977).** *Pédogénèse et classification*. Paris: Masson S.A.
- Dumont, M. G. & Murrell, J. C. (2005).** Stable isotope probing - linking microbial identity to function. *Nat Rev Microbiol* **3**, 499-504.
- Dumont, M. G., Radajewski, S. M., Miguez, C. B., McDonald, I. R. & Murrell, J. C. (2006).** Identification of a complete methane monooxygenase operon from soil by combining stable isotope probing and metagenomic analysis. *Environ Microbiol* **8**, 1240-1250.
- Edwards, R. A. & Rohwer, F. (2005).** Viral metagenomics. *Nat Rev Microbiol* **3**, 504-510.

- Eisen, J. A. (2007).** Environmental Shotgun Sequencing: Its Potential and Challenges for Studying the Hidden World of Microbes. *PLoS Biol* **5**, e82.
- El Karoui, M., Biaudet, V., Schbath, S. & Gruss, A. (1999).** Characteristics of Chi distribution on different bacterial genomes. *Res Microbiol* **150**, 579-587.
- Ellermeier, C. D., Hobbs, E. C., Gonzalez-Pastor, J. E. & Losick, R. (2006).** A three-protein signaling pathway governing immunity to a bacterial cannibalism toxin. *Cell* **124**, 549-559.
- Elliott, T. (1992).** A method for constructing single-copy lac fusions in *Salmonella typhimurium* and its application to the *hemA-prfA* operon. *J Bacteriol* **174**, 245-253.
- Engelberg-Kulka, H., Hazan, R. & Amitai, S. (2005).** *mazEF*: a chromosomal toxin-antitoxin module that triggers programmed cell death in bacteria. *J Cell Sci* **118**, 4327-4332.
- Entcheva, P., Liebl, W., Johann, A., Hartsch, T. & Streit, W. R. (2001).** Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Appl Environ Microbiol* **67**, 89-99.
- Ferrer, M., Chernikova, T. N., Yakimov, M. M., Golyshin, P. N. & Timmis, K. N. (2003).** Chaperonins govern growth of *Escherichia coli* at low temperatures. *Nat Biotechnol* **21**, 1266-1267.
- Fleischmann, R. D., Adams, M. D., White, O. & other authors (1995).** Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496-512.
- Fortin, N., Beaumier, D., Lee, K. & Greer, C. W. (2004).** Soil washing improves the recovery of total community DNA from polluted and high organic content sediments. *J Microbiol Methods* **56**, 181-191.
- Foster, R. C. (1988).** Microenvironments of soil microorganisms. *Biology and Fertility of Soils* **6**, 189-203.
- Frey, J., Bagdasarian, M. M. & Bagdasarian, M. (1992).** Replication and copy number control of the broad-host-range plasmid RSF1010. *Gene* **113**, 101-106.
- Frostegard, A., Courtois, S., Ramisse, V., Clerc, S., Bernillon, D., Le Gall, F., Jeannin, P., Nesme, X. & Simonet, P. (1999).** Quantification of bias related to the extraction of DNA directly from soils. *Appl Environ Microbiol* **65**, 5409-5420.
- Furukawa, K. (2003).** 'Super bugs' for bioremediation. *Trends Biotechnol* **21**, 187-190.
- Gabor, E. M., de Vries, E. J. & Janssen, D. B. (2003).** Efficient recovery of environmental DNA for expression cloning by indirect extraction methods. *FEMS Microbiology Ecology* **44**, 153-163.
- Gabor, E. M., Alkema, W. B. & Janssen, D. B. (2004a).** Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ Microbiol* **6**, 879-886.

- Gabor, E. M., de Vries, E. J. & Janssen, D. B. (2004b).** Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. *Environ Microbiol* **6**, 948-958.
- Gimenez, D., Perfect, E., Rawls, W. J. & Pachepsky, Y. (1997).** Fractal models for predicting soil hydraulic properties: a review. *Engineering Geology* **48**, 161-183.
- Glacock, C. B. & Weickert, M. J. (1998).** Using chromosomal lacIQ1 to control expression of genes on high-copy-number plasmids in *Escherichia coli*. *Gene* **223**, 221-231.
- Gobat, J., Aragno, M. & Matthey, W. (2003).** *Le sol vivant*, second edn. Lausanne: Presses polytechniques et universitaires romandes.
- Gonod, L. V., Chenu, C. & Soulas, G. (2003).** Spatial variability of 2,4-dichlorophenoxyacetic acid (2,4-D) mineralisation potential at a millimetre scale in soil. *Soil Biology and Biochemistry* **35**, 373-382.
- Griffiths, A. J. F., Gelbart, W. M., Miller, J. H. & Lewontin, R. C. (1999).** *Modern genetic analysis*. New York: W.H. Freeman and company.
- Gros, R., Jocteur Monrozier, L. & Faivre, P. (2006).** Does disturbance and restoration of alpine grassland soils affect the genetic structure and diversity of bacterial and N<sub>2</sub>-fixing populations? *Environ Microbiol* **8**, 1889-1901.
- Grundmann, G. L., Dechesne, A., Bartoli, F., Flandrois, J. P., Chassé, J. L. & Kizungu, R. (2001).** Spatial modeling of nitrifiers microhabitats in soil. *Soil Science Society of America Journal* **65**, 1709-1716.
- Grundmann, G. L. (2004).** Spatial scales of soil bacterial diversity - the size of a clone. *FEMS Microbiology Ecology* **48**, 119-127.
- Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. (1995).** Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177**, 4121-4130.
- Handelsman, J. (2004).** Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* **68**, 669-685.
- Harms, K., de Vries, J. & Wackernagel, W. (2007).** A double kill gene cassette for the positive selection of transforming non-selective DNA segments in *Acinetobacter baylyi* BD413. *J Microbiol Methods* **69**, 107-115.
- Heller, K. J. (2001).** Probiotic bacteria in fermented foods: product characteristics and starter organisms. *Am J Clin Nutr* **73**, 374S-379S.
- Hery, M., Herrera, A., Vogel, T. M., Normand, P. & Navarro, E. (2005a).** Effect of carbon and nitrogen input on the bacterial community structure of Neocaledonian nickel mine spoils. *FEMS Microbiol Ecol* **51**, 333-340.

- Hery, M., Philippot, L., Meriaux, E., Poly, F., Le Roux, X. & Navarro, E. (2005b).** Nickel mine spoils revegetation attempts: effect of pioneer plants on two functional bacterial communities involved in the N-cycle. *Environ Microbiol* **7**, 486-498.
- Howeler, M., Ghiorse, W. C. & Walker, L. P. (2003).** A quantitative analysis of DNA extraction and purification from compost. *Journal of Microbiological Methods* **54**, 37-45.
- Hugenholtz, P., Goebel, B. M. & Pace, N. R. (1998).** Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* **180**, 4765-4774.
- Hugenholtz, P. (2002).** Exploring prokaryotic diversity in the genomic era. *Genome Biol* **3**, REVIEWS0003.
- Hughes, E. J. & Bayly, R. C. (1983).** Control of catechol meta-cleavage pathway in *Alcaligenes eutrophus*. *J Bacteriol* **154**, 1363-1370.
- Imshenetsky, A. A., Lysenko, S. V. & Kazakov, G. A. (1978).** Upper boundary of the biosphere. *Appl Environ Microbiol* **35**, 1-5.
- Izquierdo, J. A. & Nusslein, K. (2006).** Distribution of extensive nifH gene diversity across physical soil microenvironments. *Microb Ecol* **51**, 441-452.
- Jakes, K., Zinder, N. D. & Boon, T. (1974).** Purification and properties of colicin E3 immunity protein. *J Biol Chem* **249**, 438-444.
- Jakes, K. S. & Zinder, N. D. (1974).** Highly purified colicin E3 contains immunity protein. *Proc Natl Acad Sci U S A* **71**, 3380-3384.
- Janssen, P. H., Yates, P. S., Grinton, B. E., Taylor, P. M. & Sait, M. (2002).** Improved culturability of soil bacteria and isolation in pure culture of novel members of the divisions Acidobacteria, Actinobacteria, Proteobacteria, and Verrucomicrobia. *Appl Environ Microbiol* **68**, 2391-2396.
- Jensen, E. C., Schrader, H. S., Rieland, B., Thompson, T. L., Lee, K. W., Nickerson, K. W. & Kokjohn, T. A. (1998).** Prevalence of broad-host-range lytic bacteriophages of *Sphaerotilus natans*, *Escherichia coli*, and *Pseudomonas aeruginosa*. *Appl Environ Microbiol* **64**, 575-580.
- Jensen, L. B., Ramos, J. L., Kaneva, Z. & Molin, S. (1993).** A substrate-dependent biological containment system for *Pseudomonas putida* based on the *Escherichia coli* *gef* gene. *Appl Environ Microbiol* **59**, 3713-3717.
- Jiang, D., Huang, Q., Cai, P., Rong, X. & Chen, W. (2007).** Adsorption of *Pseudomonas putida* on clay minerals and iron oxide. *Colloids Surf B Biointerfaces* **54**, 217-221.
- Juni, E. & Janik, A. (1969).** Transformation of *Acinetobacter calco-aceticus* (Bacterium anitratum). *J Bacteriol* **98**, 281-288.
- Kannan, N., Taylor, S. S., Zhai, Y., Venter, J. C. & Manning, G. (2007).** Structural and Functional Diversity of the Microbial Kinome. *PLoS Biol* **5**, e17.



- Kauffmann, I. M., Schmitt, J. & Schmid, R. D. (2004).** DNA isolation from soil samples for cloning in different hosts. *Appl Microbiol Biotechnol* **64**, 665-670.
- Kay, E., Vogel, T. M., Bertolla, F., Nalin, R. & Simonet, P. (2002).** In situ transfer of antibiotic resistance genes from transgenic (transplastomic) tobacco plants to bacteria. *Appl Environ Microbiol* **68**, 3345-3351.
- Kirk, J. L., Beaudette, L. A., Hart, M., Moutoglis, P., Klironomos, J. N., Lee, H. & Trevors, J. T. (2004).** Methods of studying soil microbial diversity. *J Microbiol Methods* **58**, 169-188.
- Knietsch, A., Waschkowitz, T., Bowien, S., Henne, A. & Daniel, R. (2003).** Construction and screening of metagenomic libraries derived from enrichment cultures: generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*. *Appl Environ Microbiol* **69**, 1408-1416.
- Knudsen, S., Saadbye, P., Hansen, L. H., Collier, A., Jacobsen, B. L., Schlundt, J. & Karlstrom, O. H. (1995).** Development and testing of improved suicide functions for biological containment of bacteria. *Appl Environ Microbiol* **61**, 985-991.
- Knudsen, S. M. & Karlstrom, O. H. (1991).** Development of efficient suicide mechanisms for biological containment of bacteria. *Appl Environ Microbiol* **57**, 85-92.
- Lakay, F. M., Botha, A. & Prior, B. A. (2007).** Comparative analysis of environmental DNA extraction and purification methods from different humic acid-rich soils. *J Appl Microbiol* **102**, 265-273.
- Lanzer, M. & Bujard, H. (1988).** Promoters largely determine the efficiency of repressor action. *Proc Natl Acad Sci U S A* **85**, 8973-8977.
- Leff, L. G., Dana, J. R., McArthur, J. V. & Shimkets, L. J. (1995).** Comparison of methods of DNA extraction from stream sediments. *Appl Environ Microbiol* **61**, 1141-1143.
- Lefrancois, J. & Sicard, A. M. (1997).** Electrotransformation of *Streptococcus pneumoniae*: evidence for restriction of DNA on entry. *Microbiology* **143** ( Pt 2), 523-526.
- Leveau, J. H. & Lindow, S. E. (2001).** Predictive and interpretive simulation of green fluorescent protein expression in reporter bacteria. *J Bacteriol* **183**, 6752-6762.
- Liles, M. R., Manske, B. F., Bintrim, S. B., Handelsman, J. & Goodman, R. M. (2003).** A census of rRNA genes and linked genomic sequences within a soil metagenomic library. *Appl Environ Microbiol* **69**, 2684-2691.
- Lilley, A. K. & Bailey, M. J. (1997).** The Acquisition of Indigenous Plasmids by a Genetically Marked *Pseudomonas* Population Colonizing the Sugar Beet Phytosphere Is Related to Local Environmental Conditions. *Appl Environ Microbiol* **63**, 1577-1583.
- Lindahl, V. & Bakken, L. R. (1995).** Evaluation of methods for extraction of bacteria from soil. *FEMS Microbiology Ecology* **16**, 135-142.

- Lindahl, V. (1996).** Improved soil dispersion procedures for total bacterial counts, extraction of indigenous bacteria and cell survival. *Journal of Microbiological Methods* **25**, 279-286.
- Liu, B., Zhang, F., Feng, X., Liu, Y., Yan, X., Zhang, X., Wang, L. & Zhao, L. (2006).** Thauera and Azoarcus as functionally important genera in a denitrifying quinoline-removal bioreactor as revealed by microbial community structure comparison. *FEMS Microbiol Ecol* **55**, 274-286.
- Liu, Y., Yang, S. F., Li, Y., Xu, H., Qin, L. & Tay, J. H. (2004).** The influence of cell and substratum surface hydrophobicities on microbial attachment. *J Biotechnol* **110**, 251-256.
- Lofblom, J., Kronqvist, N., Uhlen, M., Stahl, S. & Wernerus, H. (2007).** Optimization of electroporation-mediated transformation: Staphylococcus carnosus as model organism. *J Appl Microbiol* **102**, 736-747.
- Lombard, N., Bailly, J., Marmeisse, R. & Simonet, P. (2006).** La métagénomique des communautés microbiennes. *Biofutur* **268**, 24-27.
- Lorenz, M. G., Gerjets, D. & Wackernagel, W. (1991).** Release of transforming plasmid and chromosomal DNA from two cultured soil bacteria. *Arch Microbiol* **156**, 319-326.
- Lorenz, M. G. & Wackernagel, W. (1994).** Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol Rev* **58**, 563-602.
- Lucy, M., Reed, E. & Glick, B. R. (2004).** Applications of free living plant growth-promoting rhizobacteria. *Antonie Van Leeuwenhoek* **86**, 1-25.
- Luna, G. M., Dell'Anno, A. & Danovaro, R. (2006).** DNA extraction procedure: a critical issue for bacterial diversity assessment in marine sediments. *Environ Microbiol* **8**, 308-320.
- Lutz, R. & Bujard, H. (1997).** Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* **25**, 1203-1210.
- MacNeil, I. A., Tiong, C. L., Minor, C. & other authors (2001).** Expression and isolation of antimicrobial small molecules from soil DNA libraries. *J Mol Microbiol Biotechnol* **3**, 301-308.
- Maraha, N., Backman, A. & Jansson, J. K. (2004).** Monitoring physiological status of GFP-tagged Pseudomonas fluorescens SBW25 under different nutrient conditions and in soil by flow cytometry. *FEMS Microbiol Ecol* **51**, 123-132.
- Marsh, P. & Wellington, E. M. H. (1994).** Phage-host interactions in soil. *FEMS Microbiology Ecology* **15**, 99-107.
- Martinez, A., Kolvek, S. J., Yip, C. L., Hopke, J., Brown, K. A., MacNeil, I. A. & Osburne, M. S. (2004).** Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting

heterologous natural products in multiple expression hosts. *Appl Environ Microbiol* **70**, 2452-2463.

**Masaki, H. & Ohta, T. (1982).** A plasmid region encoding the active fragment and the inhibitor protein of colicin E3--CA38. *FEBS Lett* **149**, 129-132.

**Matic, I., Taddei, F. & Radman, M. (1996).** Genetic barriers among bacteria. *Trends Microbiol* **4**, 69-72.

**Metsa-Ketela, M., Halo, L., Munukka, E., Hakala, J., Mantsala, P. & Ylihonko, K. (2002).** Molecular evolution of aromatic polyketides and comparative sequence analysis of polyketide ketosynthase and 16S ribosomal DNA genes from various streptomyces species. *Appl Environ Microbiol* **68**, 4472-4479.

**Michel, B. (2005).** After 30 years of study, the bacterial SOS response still surprises us. *PLoS Biol* **3**, e255.

**Miki, B. & McHugh, S. (2004).** Selectable marker genes in transgenic plants: applications, alternatives and biosafety. *J Biotechnol* **107**, 193-232.

**Millan, H., Gonzalez-Posada, M., Morilla, A. A. & Perez, E. (2007).** Self-similar organization of Vertisol microstructure: A pore-solid fractal interpretation. *Geoderma* **138**, 185-190.

**Molin, S., Boe, L., Jensen, L. B., Kristensen, C. S., Givskov, M., Ramos, J. L. & Bej, A. K. (1993).** Suicidal genetic elements and their use in biological containment of bacteria. *Annu Rev Microbiol* **47**, 139-166.

**Molina, L., Ramos, C., Ronchel, M. C., Molin, S. & Ramos, J. L. (1998).** Construction of an efficient biologically contained pseudomonas putida strain and its survival in outdoor assays. *Appl Environ Microbiol* **64**, 2072-2078.

**Moller, S., Sternberg, C., Andersen, J. B., Christensen, B. B., Ramos, J. L., Givskov, M. & Molin, S. (1998).** In situ gene expression in mixed-culture biofilms: evidence of metabolic interactions between community members. *Appl Environ Microbiol* **64**, 721-732.

**Morales, V. M., Backman, A. & Bagdasarian, M. (1991).** A series of wide-host-range low-copy-number vectors that allow direct screening for recombinants. *Gene* **97**, 39-47.

**Muller, A. K., Westergaard, K., Christensen, S. & Sorensen, S. J. (2002).** The diversity and function of soil microbial communities exposed to different disturbances. *Microb Ecol* **44**, 49-58.

**Mummey, D., Holben, W., Six, J. & Stahl, P. (2006).** Spatial stratification of soil bacterial populations in aggregates of diverse soils. *Microb Ecol* **51**, 404-411.

**Mummey, D. L. & Stahl, P. D. (2004).** Analysis of soil whole- and inner-microaggregate bacterial communities. *Microb Ecol* **48**, 41-50.

- Muniyappa, K., Shaner, S. L., Tsang, S. S. & Radding, C. M. (1984).** Mechanism of the concerted action of recA protein and helix-destabilizing proteins in homologous recombination. *Proc Natl Acad Sci U S A* **81**, 2757-2761.
- Munthali, M. T., Timmis, K. N. & Diaz, E. (1996).** Use of Colicin E3 for Biological Containment of Microorganisms. *Appl Environ Microbiol* **62**, 1805-1807.
- Murphy, K. C. (1991).** Lambda Gam protein inhibits the helicase and chi-stimulated recombination activities of Escherichia coli RecBCD enzyme. *J Bacteriol* **173**, 5808-5821.
- Murphy, K. C. (1998).** Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli. *J Bacteriol* **180**, 2063-2071.
- Nannipieri, P., Ascher, J., Ceccherini, M. T., Landi, L., Pietramellara, G. & Renella, G. (2003).** Microbial diversity and soil functions. *European Journal of Soil Science* **54**, 655-670.
- Newman, J. R. & Fuqua, C. (1999).** Broad-host-range expression vectors that carry the L-arabinose-inducible Escherichia coli araBAD promoter and the araC regulator. *Gene* **227**, 197-203.
- Nielsen, K. M., Bones, A. M. & Van Elsas, J. D. (1997a).** Induced Natural Transformation of Acinetobacter calcoaceticus in Soil Microcosms. *Appl Environ Microbiol* **63**, 3972-3977.
- Nielsen, K. M., van Weerelt, M. D., Berg, T. N., Bones, A. M., Hagler, A. N. & van Elsas, J. D. (1997b).** Natural transformation and availability of transforming DNA to Acinetobacter calcoaceticus in soil microcosms. *Appl Environ Microbiol* **63**, 1945-1952.
- Nielsen, K. M., Bones, A. M., Smalla, K. & van Elsas, J. D. (1998).** Horizontal gene transfer from transgenic plants to terrestrial bacteria--a rare event? *FEMS Microbiol Rev* **22**, 79-103.
- Nielsen, K. M., Smalla, K. & van Elsas, J. D. (2000).** Natural transformation of Acinetobacter sp. strain BD413 with cell lysates of Acinetobacter sp., Pseudomonas fluorescens, and Burkholderia cepacia in soil microcosms. *Appl Environ Microbiol* **66**, 206-212.
- Nordstrom, K. & Austin, S. J. (1989).** Mechanisms that contribute to the stable segregation of plasmids. *Annu Rev Genet* **23**, 37-69.
- Nunan, N., Wu, K., Young, I. M., Crawford, J. W. & Ritz, K. (2002).** In situ spatial patterns of soil bacterial populations, mapped at multiple scales, in an arable soil. *Microb Ecol* **44**, 296-305.
- Nunan, N., Wu, K., Young, I. M., Crawford, J. W. & Ritz, K. (2003).** Spatial distribution of bacterial communities and their relationships with the micro-architecture of soil. *FEMS Microbiology Ecology* **44**, 203-215.
- Ogram, A., Saylor, G. S. & Barkay, T. (1987).** The extraction and purification of microbial DNA from sediments. *Journal of Microbiological Methods* **7**, 57-66.

- Ohashi-Kunihiro, S., Hagiwara, H., Yohda, M., Masaki, H. & Machida, M. (2006).** Construction of a positive selection marker by a lethal gene with the amber stop codon(s) regulator. *Biosci Biotechnol Biochem* **70**, 119-125.
- Paget, E. & Simonet, P. (1994).** On the track of natural transformation in soil. *FEMS Microbiology Ecology* **15**, 109-117.
- Pallud, C., Dechesne, A., Gaudet, J. P., Debouzie, D. & Grundmann, G. L. (2004).** Modification of spatial distribution of 2,4-dichlorophenoxyacetic acid degrader microhabitats during growth in soil columns. *Appl Environ Microbiol* **70**, 2709-2716.
- Palmen, R. & Hellingwerf, K. J. (1997).** Uptake and processing of DNA by *Acinetobacter calcoaceticus*--a review. *Gene* **192**, 179-190.
- Parent, M. E. & Velegol, D. (2004).** E. coli adhesion to silica in the presence of humic acid. *Colloids Surf B Biointerfaces* **39**, 45-51.
- Parish, T. & Stoker, N. G. (2000).** glnE is an essential gene in *Mycobacterium tuberculosis*. *J Bacteriol* **182**, 5715-5720.
- Philippot, L., Piutti, S., Martin-Laurent, F., Hallet, S. & Germon, J. C. (2002).** Molecular analysis of the nitrate-reducing community from unplanted and maize-planted soils. *Appl Environ Microbiol* **68**, 6121-6128.
- Poteete, A. R., Fenton, A. C. & Murphy, K. C. (1988).** Modulation of *Escherichia coli* RecBCD activity by the bacteriophage lambda Gam and P22 Abc functions. *J Bacteriol* **170**, 2012-2021.
- Poteete, A. R. (2001).** What makes the bacteriophage lambda Red system useful for genetic engineering: molecular mechanism and biological function. *FEMS Microbiol Lett* **201**, 9-14.
- Raaijmakers, J. M., Vlami, M. & de Souza, J. T. (2002).** Antibiotic production by bacterial biocontrol agents. *Antonie Van Leeuwenhoek* **81**, 537-547.
- Ranjard, L., Poly, F., Combrisson, J., Richaume, A. & Nazaret, S. (1998).** A single procedure to recover DNA from the surface or inside aggregates and in various size fractions of soil suitable for PCR-based assays of bacterial communities. *European Journal of Soil Biology* **34**, 89-97.
- Ranjard, L., Poly, F. & Nazaret, S. (2000).** Monitoring complex bacterial communities using culture-independent molecular techniques: application to soil environment. *Res Microbiol* **151**, 167-177.
- Ranjard, L. & Richaume, A. (2001).** Quantitative and qualitative microscale distribution of bacteria in soil. *Res Microbiol* **152**, 707-716.
- Rappoldt, C. & Crawford, J. W. (1999).** The distribution of anoxic volume in a fractal model of soil. *Geoderma* **88**, 329-347.

- Recorbet, G., Robert, C., Givaudan, A., Kudla, B., Normand, P. & Faurie, G. (1993).** Conditional suicide system of *Escherichia coli* released into soil that uses the *Bacillus subtilis* *sacB* gene. *Appl Environ Microbiol* **59**, 1361-1366.
- Richaume, A., Angle, J. S. & Sadowsky, M. J. (1989).** Influence of soil variables on in situ plasmid transfer from *Escherichia coli* to *Rhizobium fredii*. *Appl Environ Microbiol* **55**, 1730-1734.
- Richter, M., Milling, A. & Smalla, K. (2003).** International symposium of structure and function of soil microbiota. In *Monitoring of horizontal gene transfer from transplastomic plants to Acinetobacter sp (pFG4) and screening for the capability of natural transformation within bacteria of rhizosphere and soil*. Germany.
- Riesenfeld, C. S., Goodman, R. M. & Handelsman, J. (2004a).** Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environ Microbiol* **6**, 981-989.
- Riesenfeld, C. S., Schloss, P. D. & Handelsman, J. (2004b).** Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet* **38**, 525-552.
- Riley, M. A. & Wertz, J. E. (2002).** Bacteriocins: evolution, ecology, and application. *Annu Rev Microbiol* **56**, 117-137.
- Robe, P., Nalin, R., Capellano, C., Vogel, T. M. & Simonet, P. (2003).** Extraction of DNA from soil. *European Journal of Soil Biology* **39**, 183-190.
- Romanowski, G., Lorenz, M. G., Sayler, G. & Wackernagel, W. (1992).** Persistence of Free Plasmid DNA in Soil Monitored by Various Methods, Including a Transformation Assay. *Appl Environ Microbiol* **58**, 3012-3019.
- Romanowski, G., Lorenz, M. G. & Wackernagel, W. (1993).** Plasmid DNA in a groundwater aquifer microcosm--adsorption, DNAase resistance and natural genetic transformation of *Bacillus subtilis*. *Mol Ecol* **2**, 171-181.
- Ronchel, M. C., Ramos, C., Jensen, L. B., Molin, S. & Ramos, J. L. (1995).** Construction and behavior of biologically contained bacteria for environmental applications in bioremediation. *Appl Environ Microbiol* **61**, 2990-2994.
- Ronchel, M. C. & Ramos, J. L. (2001).** Dual system to reinforce biological containment of recombinant bacteria designed for rhizoremediation. *Appl Environ Microbiol* **67**, 2649-2656.
- Rondon, M. R., August, P. R., Bettermann, A. D. & other authors (2000).** Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* **66**, 2541-2547.
- Rosenberg, M. (2006).** Microbial adhesion to hydrocarbons: twenty-five years of doing MATH. *FEMS Microbiol Lett* **262**, 129-134.
- Saida, F., Uzan, M., Odaert, B. & Bontems, F. (2006).** Expression of highly toxic genes in *E. coli*: special strategies and genetic tools. *Curr Protein Pept Sci* **7**, 47-56.

- Sanguin, H., Remenant, B., Dechesne, A., Thioulouse, J., Vogel, T. M., Nesme, X., Moenne-Loccoz, Y. & Grundmann, G. L. (2006).** Potential of a 16S rRNA-based taxonomic microarray for analyzing the rhizosphere effects of maize on *Agrobacterium* spp. and bacterial communities. *Appl Environ Microbiol* **72**, 4302-4312.
- Sat, B., Reches, M. & Engelberg-Kulka, H. (2003).** The *Escherichia coli* mazEF suicide module mediates thymineless death. *J Bacteriol* **185**, 1803-1807.
- Schink, B. (2002).** Synergistic interactions in the microbial world. *Antonie Van Leeuwenhoek* **81**, 257-261.
- Schloss, P. D. & Handelsman, J. (2003).** Biotechnological prospects from metagenomics. *Curr Opin Biotechnol* **14**, 303-310.
- Schloss, P. D. & Handelsman, J. (2004).** Status of the microbial census. *Microbiol Mol Biol Rev* **68**, 686-691.
- Schurgers, G., Dorsch, P., Bakken, L., Leffelaar, P. & Haugen, L. E. (2006).** Modelling soil anaerobiosis from water retention characteristics and soil respiration. *Soil Biology and Biochemistry* **38**, 2637-2644.
- Sengelov, G. & Sorensen, S. J. (1998).** Methods for detection of conjugative plasmid transfer in aquatic environments. *Curr Microbiol* **37**, 274-280.
- Senior, B. W. & Holland, I. B. (1971).** Effect of colicin E3 upon the 30S ribosomal subunit of *Escherichia coli*. *Proc Natl Acad Sci U S A* **68**, 959-963.
- Setlow, J. K., Randesi, M., Adams, J. G., Setlow, B. & Setlow, P. (1992).** Mutation and killing of *Escherichia coli* expressing a cloned *Bacillus subtilis* gene whose product alters DNA conformation. *J Bacteriol* **174**, 2943-2950.
- Shen, P. & Huang, H. V. (1986).** Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. *Genetics* **112**, 441-457.
- Siegele, D. A. & Hu, J. C. (1997).** Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. *Proc Natl Acad Sci U S A* **94**, 8168-8172.
- Simon, R., Priefer, U. & Puehler, A. (1983).** A broad host range mobilization system for in vivo genetic engineering transposon mutagenesis in Gram-negative bacteria. *Biotechnology* **1**, 784-791.
- Simonet, P., Capellano, A., Navarro, E., Bardin, R. & Moiroud, A. (1984).** An improved method for lysis of *Frankia* with achromopeptidase allows detection of new plasmids. *Canadian Journal of Microbiology* **30**, 1292-1295.
- Simons, R. W., Houman, F. & Kleckner, N. (1987).** Improved single and multicopy lac-based cloning vectors for protein and operon fusions. *Gene* **53**, 85-96.

- Smith, Z., McCaig, A. E., Stephen, J. R., Embley, T. M. & Prosser, J. I. (2001).** Species Diversity of Uncultured and Cultured Populations of Soil and Marine Ammonia Oxidizing Bacteria. *Microb Ecol* **42**, 228-237.
- Sorokin, D. Y., Tourova, T. P., Lysenko, A. M. & Muyzer, G. (2006).** Diversity of culturable halophilic sulfur-oxidizing bacteria in hypersaline habitats. *Microbiology* **152**, 3013-3023.
- Sosio, M., Giusino, F., Cappellano, C., Bossi, E., Puglia, A. M. & Donadio, S. (2000).** Artificial chromosomes for antibiotic-producing actinomycetes. *Nat Biotechnol* **18**, 343-345.
- Steffan, R. J., Goksoyr, J., Bej, A. K. & Atlas, R. M. (1988).** Recovery of DNA from soils and sediments. *Appl Environ Microbiol* **54**, 2908-2915.
- Stein, J. L., Marsh, T. L., Wu, K. Y., Shizuya, H. & DeLong, E. F. (1996).** Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J Bacteriol* **178**, 591-599.
- Stenstrom, T. A. (1989).** Bacterial hydrophobicity, an overall parameter for the measurement of adhesion potential to soil particles. *Appl Environ Microbiol* **55**, 142-147.
- Stetter, K. O. (1999).** Extremophiles and their adaptation to hot environments. *FEBS Lett* **452**, 22-25.
- Stewart, G. J., Carlson, C. A. & Ingraham, J. L. (1983).** Evidence for an active role of donor cells in natural transformation of *Pseudomonas stutzeri*. *J Bacteriol* **156**, 30-35.
- Strevett, K. A. & Chen, G. (2003).** Microbial surface thermodynamics and applications. *Res Microbiol* **154**, 329-335.
- Szafranski, P., Mello, C. M., Sano, T., Smith, C. L., Kaplan, D. L. & Cantor, C. R. (1997).** A new approach for containment of microorganisms: dual control of streptavidin expression by antisense RNA and the T7 transcription system. *Proc Natl Acad Sci U S A* **94**, 1059-1063.
- Szipirer, C., Top, E., Couturier, M. & Mergeay, M. (1999).** Retrotransfer or gene capture: a feature of conjugative plasmids, with ecological and evolutionary significance. *Microbiology* **145** ( Pt 12), 3321-3329.
- Thomas, C. M. & Nielsen, K. M. (2005).** Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* **3**, 711-721.
- Thomason, L. C., Court, D. L., Bubunencko, M., Costantino, N., Wilson, H., Datta, S. & Oppenheim, A. (2005).** *Recombineering: genetic engineering in bacteria using homologous recombination* Hoboken: John Wiley & Sons Inc.
- Tock, M. R. & Dryden, D. T. (2005).** The biology of restriction and anti-restriction. *Curr Opin Microbiol* **8**, 466-472.



- Torres, B., Jaenecke, S., Timmis, K. N., Garcia, J. L. & Diaz, E. (2000).** A gene containment strategy based on a restriction-modification system. *Environ Microbiol* **2**, 555-563.
- Torres, B., Jaenecke, S., Timmis, K. N., Garcia, J. L. & Diaz, E. (2003).** A dual lethal system to enhance containment of recombinant micro-organisms. *Microbiology* **149**, 3595-3601.
- Torsvik, V., Goksoyr, J. & Daae, F. L. (1990).** High diversity in DNA of soil bacteria. *Appl Environ Microbiol* **56**, 782-787.
- Torsvik, V., Ovreas, L. & Thingstad, T. F. (2002).** Prokaryotic diversity--magnitude, dynamics, and controlling factors. *Science* **296**, 1064-1066.
- Treusch, A. H., Kletzin, A., Raddatz, G., Ochsenreiter, T., Quaiser, A., Meurer, G., Schuster, S. C. & Schleper, C. (2004).** Characterization of large-insert DNA libraries from soil for environmental genomic studies of Archaea. *Environ Microbiol* **6**, 970-980.
- Tringe, S. G., von Mering, C., Kobayashi, A. & other authors (2005).** Comparative metagenomics of microbial communities. *Science* **308**, 554-557.
- Tyler, I. M., Morris, P. A., Thorne, A. M., Sheppard, S., Smithies, R. H., Riganti, A., Doyle, M. G., and Hocking, R. M. (2004).** The revised GSWA rock classification scheme. In *Geological survey of western australia annual review*, pp. 88-94.
- Tyson, G. W., Chapman, J., Hugenholtz, P. & other authors (2004).** Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37-43.
- Uchiyama, T., Abe, T., Ikemura, T. & Watanabe, K. (2005).** Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nat Biotechnol* **23**, 88-93.
- van der Rest, M. E., Lange, C. & Molenaar, D. (1999).** A heat shock following electroporation induces highly efficient transformation of *Corynebacterium glutamicum* with xenogeneic plasmid DNA. *Appl Microbiol Biotechnol* **52**, 541-545.
- van Dillewijn, P., Vilchez, S., Paz, J. A. & Ramos, J. L. (2004).** Plant-dependent active biological containment system for recombinant rhizobacteria. *Environ Microbiol* **6**, 88-92.
- van Loosdrecht, M. C., Lyklema, J., Norde, W., Schraa, G. & Zehnder, A. J. (1987).** The role of bacterial cell wall hydrophobicity in adhesion. *Appl Environ Microbiol* **53**, 1893-1897.
- Vaneechoutte, M., Young, D. M., Ornston, L. N., De Baere, T., Nemec, A., Van Der Reijden, T., Carr, E., Tjernberg, I. & Dijkshoorn, L. (2006).** Naturally transformable *Acinetobacter* sp. strain ADP1 belongs to the newly described species *Acinetobacter baylyi*. *Appl Environ Microbiol* **72**, 932-936.
- Venter, J. C., Remington, K., Heidelberg, J. F. & other authors (2004).** Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66-74.

- Vernet, T., Lau, P. C., Narang, S. A. & Visentin, L. P. (1985).** A direct-selection vector derived from pColE3-CA38 and adapted for foreign gene expression. *Gene* **34**, 87-93.
- Vincent, W. F., Mueller, D. R. & Bonilla, S. (2004).** Ecosystems on ice: the microbial ecology of Markham Ice Shelf in the high Arctic. *Cryobiology* **48**, 103-112.
- Vogel, H. J. & Roth, K. (2001).** Quantitative morphology and network representation of soil pore structure. *Advances in Water Resources* **24**, 233-242.
- von Wintzingerode, F., Gobel, U. B. & Stackebrandt, E. (1997).** Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* **21**, 213-229.
- Vulic, M., Lenski, R. E. & Radman, M. (1999).** Mutation, recombination, and incipient speciation of bacteria in the laboratory. *Proc Natl Acad Sci U S A* **96**, 7348-7351.
- Walker, S. G., Flemming, C. A., Ferris, F. G., Beveridge, T. J. & Bailey, G. W. (1989).** Physicochemical interaction of *Escherichia coli* cell envelopes and *Bacillus subtilis* cell walls with two clays and ability of the composite to immobilize heavy metals from solution. *Appl Environ Microbiol* **55**, 2976-2984.
- Wang, G. Y., Graziani, E., Waters, B. & other authors (2000).** Novel natural products from soil DNA libraries in a streptomycete host. *Org Lett* **2**, 2401-2404.
- Wawrik, B., Kerkhof, L., Zylstra, G. J. & Kukor, J. J. (2005).** Identification of unique type II polyketide synthase genes in soil. *Appl Environ Microbiol* **71**, 2232-2238.
- Weaver, J. C. (1995).** Electroporation theory. Concepts and mechanisms. *Methods Mol Biol* **55**, 3-28.
- Weinbauer, M. G. (2004).** Ecology of prokaryotic viruses. *FEMS Microbiol Rev* **28**, 127-181.
- Wellington, E. M., Berry, A. & Krsek, M. (2003).** Resolving functional diversity in relation to microbial community structure in soil: exploiting genomics and stable isotope probing. *Curr Opin Microbiol* **6**, 295-301.
- White, R. E. (2006).** *Principles and practice of soil science. The soil as a natural resource*, Fourth edn: Blackwell Publishing.
- Whitman, W. B., Coleman, D. C. & Wiebe, W. J. (1998).** Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A* **95**, 6578-6583.
- Wilson, I. G. (1997).** Inhibition and facilitation of nucleic acid amplification. *Appl Environ Microbiol* **63**, 3741-3751.
- Woese, C. R. (1987).** Bacterial evolution. *Microbiol Rev* **51**, 221-271.

- Yooseph, S., Sutton, G., Rusch, D. B. & other authors (2007).** The Sorcerer II Global Ocean Sampling Expedition: Expanding the Universe of Protein Families. *PLoS Biol* **5**, e16.
- Young, I. M. & Crawford, J. W. (2004).** Interactions and self-organization in the soil-microbe complex. *Science* **304**, 1634-1637.
- Yu, D., Ellis, H. M., Lee, E. C., Jenkins, N. A., Copeland, N. G. & Court, D. L. (2000).** An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci U S A* **97**, 5978-5983.
- Zengler, K., Toledo, G., Rappe, M., Elkins, J., Mathur, E. J., Short, J. M. & Keller, M. (2002).** Cultivating the uncultured. *Proc Natl Acad Sci U S A* **99**, 15681-15686.
- Zeph, L. R., Onaga, M. A. & Stotzky, G. (1988).** Transduction of *Escherichia coli* by bacteriophage P1 in soil. *Appl Environ Microbiol* **54**, 1731-1737.
- Zhang, X., Gao, P., Chao, Q., Wang, L., Senior, E. & Zhao, L. (2004).** Microdiversity of phenol hydroxylase genes among phenol-degrading isolates of *Alcaligenes* sp. from an activated sludge system. *FEMS Microbiol Lett* **237**, 369-375.
- Zhang, X. Z., Yan, X., Cui, Z. L., Hong, Q. & Li, S. P. (2006).** mazF, a novel counter-selectable marker for unmarked chromosomal manipulation in *Bacillus subtilis*. *Nucleic Acids Res* **34**, e71.
- Zhou, J., Bruns, M. A. & Tiedje, J. M. (1996).** DNA recovery from soils of diverse composition. *Appl Environ Microbiol* **62**, 316-322.

## **ANNEXE 1 : Composition des milieux de culture et des tampons utilisés**

### Milieu Luria Bertani (LB)

Composé	Concentration initiale	Quantité	Concentration finale
Extrait de levure		5 g	0,5 %
Bacto tryptone		10 g	1 %
NaCl <sup>a</sup>		10 g	~0,17 M
Agar <sup>b</sup>		10 g	1 %
Eau		1 L	qsp

<sup>a</sup> Le milieu LBM contient moitié moins de chlorure de sodium

<sup>b</sup> L'agar est enlevé pour les milieux liquide

### Tampon colicine

Composé	Concentration initiale	Quantité	Concentration finale
NaH <sub>2</sub> PO <sub>4</sub>		1,56 g	
NaCl		58,4 g	1 M
Eau		1 L	qsp

Ajuster a pH 7 avec NaOH

### Tampon de lyse

Composé	Concentration initiale	Quantité	Concentration finale
Tris	1 M	2,5 mL	25 mM
EDTA	0,5 M	2 mL	10 mM
Saccharose		2 g	20 %
Eau		95,5 mL	Qsp 100 mL

### Solution Détergente

Composé	Concentration initiale	Quantité	Concentration finale
SDS	10 %	10 mL	1 %
NaOH	1 N	20 mL	0,2 N
Eau		70 mL	Qsp 100 mL

## **ANNEXE 2 : Protocole pour l'isolement de plasmide de *E. coli* et purification sur gradient de Chlorure de césium CsCl**

### **Matériel :**

Solution de lyse  
Solution détergente  
Solution neutralisation  
Tube polycarbonate nalgène 38 mL  
Tube ultracentrifugation polyallomère beckman 12 mL  
Ultracentrifugeuse Kontron TGA-50, rotor TST 41.14  
Réfractomètre

### **Méthode :**

#### **1. Culture bactérienne**

- Inoculer 1 colonie isolée fraîche dans 10 mL liquide LB + antibiotique. Incuber 8 h à 37°C sous agitation
- Inoculer 1/100<sup>ème</sup> préculture (5 mL) dans 500 mL LB+antibiotiques. Incuber 16 h à 37°C sous agitation
- Centrifuger la culture en 2 x 250 mL à 6 000 rpm pendant 10 min 4°C (centrifugeuse Beckman T2-21M/E rotor JA10.0 (6370g))
- Eliminer totalement le surnageant (retourner les pots sur un papier absorbant)

#### **2. Lyse cellulaire**

- Bien reprendre chaque échantillon de 250 mL avec 5 mL du tampon de lyse par pipetage/relargage. Transférer dans un tube en polycarbonate (soit 2 tubes pour les 500 mL de culture initiale).
- Ajouter quelque mg de lysosyme (pointe de spatule).
- Incuber 30 min dans la glace
- Ajouter dans chaque tube 10 mL de la solution de détergent. Homogénéiser par retournements jusqu'à l'obtention d'une solution visqueuse (6-8 fois).
- Incuber 5 min dans la glace<sup>a</sup>
- Ajouter dans chaque tube 7,5 mL de la solution de neutralisation (soit 2x7.5 mL pour les 500 mL de culture initiale). Homogénéiser par retournements (6-8 fois) jusqu'à l'obtention d'une solution limpide (non visqueuse)
- Incuber 60 min dans la glace

#### **3. Clarification du lysat**

- Centrifuger 20 min à 8 000 rpm 4°C (centrifugeuse SORVALL 5C 5B plus, rotor SS-34, 7 670 g).
- Filtrer le surnageant au travers d'une gaze afin d'éliminer tout lysat. Remise en commun des 2 sous-échantillons dans 1 tube NALGENE 38 mL

4. Précipitation de l'ADN

- Ajouter 2 volumes d'éthanol pur à -20°C ou 1 volume d'isopropanol
- Centrifuger pendant 20 min à 15 000 g (11 000 rpm sur SS-34) à 4°C pour l'éthanol ou 20°C pour l'isopropanol<sup>b</sup>
- Laver le culot d'ADN avec 10-20 mL d'éthanol 70% préalablement refroidi à -20°C (mais pas de remise en suspension)
- Centrifuger 20 min à 15 000 g à 4°C.
- Sécher le culot d'ADN 15 min environ à 37°C
- Reprendre l'ADN dans 6,5 mL de TE (Tris+EDTA 10:1) pour préparation du gradient de CsCl

5. Purification sur gradient de CsCl

- Transférer le culot repris dans le TE dans des tubes d'ultracentrifugation
- Peser exactement 6.83 g de CsCl, les ajouter dans chaque tube d'ultracentrifugation.
- Ajouter 0.33 mL de BET à 10mg.mL<sup>-1</sup> (1 g.mL<sup>-1</sup> CsCl, réfractomètre : 1,4 ; correspond à une densité de 1,707)
- Ultracentrifuger à 33 000 rpm à 15°C (25°C max) pendant 40 h
- Prélever l'ADN à l'aide d'une seringue sous UVs

6. Elimination du BET

- Mélanger de l'isopropanol et du SSC 20 X (v :v).
- Ajouter quelques mL d'isopropanol (phase supérieure) à la fraction prélevée, mélanger par retournement, le BET se dissout dans la phase organique.

<sup>a</sup> dénaturation de l'ADN, ne pas incuber davantage car risque de ne pas pouvoir renaturer l'ADN plasmidique

<sup>b</sup> risque de précipitation des sels si centrifugation à 4°C avec l'isopropanol

<sup>c</sup> le SSC sature l'isopropanol

### **ANNEXE 3 : Protocole de purification d'ADN à partir d'un gel d'agarose**

#### **Matériel :**

Cuve électrophorèse

Plaque UV

Casque de protection, gants, blouse

#### **Méthode :**

- Mélanger la solution d'ADN à du bleu de bromophénol selon un ratio en volume 1 :5
- Déposer le mélange dans un puits de gel d'agarose ne contenant pas de BET, ainsi qu'un marqueur de taille.
- Effectuer la migration dans du tampon TBE 0,5 X sous une tension de 50 V
- Placer le gel 1 min dans un bain de BET à  $1\mu\text{g.mL}^{-1}$  en concentration finale
- Laver le gel dans un bain d'eau ultra pure quelques minutes.
- Exciser le fragment d'ADN de taille attendue sous UVs
- Déposer le morceau d'agarose contenant l'ADN dans un tube eppendorf pré-taré
- Purifier l'ADN est sur colonne commerciale selon le protocole du fabricant

## **ANNEXE 4 : Protocole de déphosphorylation de l'ADN**

### Matériel :

Phosphatase alcaline shrimp SAP (Roche)  
 Tampon associé (Roche)  
 Eau UP  
 ADN  
 Etuve ou bain marie ou bain sec  
 Gants, cônes, pipettes, portoir ouvert, bloc réfrigérant

### Méthode :

#### 1. Préparation mélange réactionnel

Composants	Concentration initiale	Volume (µL)	Concentration finale
ADN		x	~500 ng
Eau		x	Qsp 25 µL
Tampon	10 X	3	1 X
Enzyme SAP	1 U.µL <sup>-1</sup>	0,25 µL	~0,01 U.µL <sup>-1a</sup>

#### 2. Réaction enzymatique

- Incuber 30 min à 37°C
- Ajouter même volume d'enzyme
- Incuber 30 min à 37°C

#### 3. Inactivation enzyme

- Inactiver enzyme à 70°C pendant 10 min
- Incuber 5 min sur glace
- Centrifuger quelques secondes pour sédimenter les gouttes condensées
- Purification ADN

#### 4. Contrôle de la déphosphorylation

par transformation de :

plasmide linéarisé déphosphorylé ligué	Essai	0
plasmide linéarisé ligué	Témoin positif	transf.
plasmide linéarisé	Témoin négatif	0

Plasmide linéarisé à bouts cohésifs	Masse (ng)	SAP(U.µL <sup>-1</sup> )	Fréquence transformation
pUC19	50	0,01	<10 <sup>-8</sup>
pUC19	50	0,1	10 <sup>-8</sup>
pMMB190	50	0,01	10 <sup>-8</sup>
pMMB190	50	0,1	10 <sup>-7</sup>

<sup>a</sup> Meilleure déphosphorylation si l'enzyme est utilisée diluée au 1/10<sup>ème</sup> et introduite dans le mélange en 2 fois (Mavingui, communication personnelle). Ce phénomène a été vérifié par un contrôle de la déphosphorylation comme indiqué à l'étape 4 avec déphosphorylation 1 étape 10 min



## **ANNEXE 5 : Protocole d'extraction de la colicine à partir de souches *E. coli* productrices**

### Matériel :

Tampon colicine  
Fioles à toxines  
Mytomicine C  
Hotte chimique

### Méthode :

1. Préculture n°1 (Vf = 5 mL)
  - inoculer 1 colonie dans tube contenant 5 mL LB
  - incuber à 37°C à 150 rpm, 1 journée
2. Préculture n°2 (Vf = 40 mL)
  - inoculer 5 mL préculture n°1 dans erlen contenant 40 mL LB,
  - incuber à 37°C à 150 rpm, 1 nuit
  - = souche en fin de phase exponentielle, début de phase stationnaire
3. Culture souche productrice (Vf = 1L)
  - inoculer 45 mL préculture n°2 dans fiole à toxine (ou erlen 2 L) contenant 1 L LB préchauffé à 37°C
  - incuber à 37°C à 150 rpm, jusqu'à DO= 0,8 à 1 (~2 h seulement)
4. Induction production de la colicine avec 1mg.L<sup>-1</sup> de mytomicine C
  - dans flacon contenant mytomicine C en poudre, injecter sous hotte chimique du méthanol pour une Cf =1 mg.mL<sup>-1</sup> (solution bleutée photosensible)
  - injecter 1 mL de mytomicine C à 1mg.mL<sup>-1</sup> dans culture
  - incuber 37°C à 150 rpm, 5 h
5. Reprise du culot de cellules dans 100 mL tampon colicine
  - Transférer la culture dans nalgène 500 mL
  - Centrifuger la culture 8 000 rpm 10 min RT (*room temperature*), rotation inclinée
  - Eliminer surnageant, renverse le nalgène sur papier absorbant pour éliminer restes
  - Reprendre le culot par 100 mL de tampon colicine
6. Lyse des cellules avec du chloroforme
  - Transférer la solution dans un erlen
  - Rajouter 1/10 volume de chloroforme sous hotte chimique
  - Incuber 37°C 150 rpm (bonne agitation nécessaire)
  - (Laisse 15 min a froid)
  - ou décanter dans un bécher à température ambiante sous hotte chimique
  - = chloroforme plus dense au fond du bécher
  - = attention, toujours utiliser du verre en présence de chloroforme, car dissous le plastique

7. Aliquote solution et stockage

- Prélèver le surnageant (tampon colicine + colicine)
- Répartir 4 mL dans tube en verre bouchonnés (bouchon backélite) sans le chloroforme
- Rajouter ~0,5 mL de chloroforme par tube
- Vortexer fortement
- Conserver à -20°C

= chloroforme au fond, puis résidus cellulaires, puis tampon+colicine en surface

8. Utilisation colicine

- décongèler et prélèver tampon+colicine+culot (mais pas le chloroforme)
- déposer 4 mL solution colicine pour 800 mL LB agar en surfusion à 50°C

## **ANNEXE 6 : Protocole de transduction d'ADN**

### **Matériel :**

Souche réceptrice  
Eau gélosée en surfusion  
LBm liquide  
Milieu MC  
CaCl<sub>2</sub> 0,2M

### **Méthode :**

1. Préparation souche donatrice
  - Preculture 1 souche dans 2 mL LB incubée à 37°C
  - Preculture 2 diluée au 1/10 LB+ goutte CaCl<sub>2</sub> 0,2 M (a). Incuber 1 h jusqu'à DO 0,2-0,4 maximum
2. Préparation phage P1
  - Mélanger 50 µL stock de phage P1 (b) à 1 mL de cellules
  - Incuber 15 min à 37°C **sans agitation**
  - Mélanger avec 4 mL agar en surfusion. Vortexer
  - Déposer la totalité sur boîte P1 non froides non séchées (c)
  - Attendre prise en masse de l'agar
  - Incuber 1 nuit à 37°C sans retourner les boîtes
  - Sortir les boîtes : milieu claire avec quelques colonies (d)
3. Transduction
  - Racler surcouches d'agar avec un rateau
  - Transvaser dans un tube (e) l'équivalent de 4 boîtes de lysat
  - Ajouter 9 mL de chloroforme. Vortexer 30 s-1 min
  - Centrifuger 12 000 rpm 10 min 4°C. Surnageant source de phage défectifs
  - Centrifuger 2 mL culture phase stationnaire de la souche receveuse 10 000 rpm 2 min pour sédimenter les cellules
  - Bien éliminer le surnageant (f) et resuspendre les cellules dans 1 mL de MC
  - Réaliser les mélanges dans tube eppendorf
    - 100 µL cellules + 100 µL phages
    - 100 µL cellules + 50 µL phages
    - 100 µL cellules + 10 µL phages
    - 100 µL phages
    - 100 µL cellules
  - Incuber 20 min 37°C
  - Etaler la totalité sur LBm +AB
  - Incuber à 37°C boîtes non retournées

Remarques :

- (a) amélioration adsorption du phage sur cellules (échange ionique entre  $Mg^{2+}$  LPS bactérien et  $Ca^{2+}$  du milieu)
- (b) stock à 4°C de phage dans surnageant LB au dessus phase chloroforme (utilisé pour stériliser et conserver phage)
- (c) boîtes doivent rester humides, donc pas plus de 2 jours sur paillasse ou conservée à 4°C dans sachet plastique
- (d) le phage P1 a lysé les cellules donatrices (sauf les résistantes). Le phage a récupéré au cours de la lyse soit son propre ADN : phage virulent (majorité des cas, réutilisé pour nouveau stock de phage) soit ADN bactérien : phage défectif (minorité, ceux qui nous intéressent)
- (e) tube résistant à la centrifugation 12 000 rpm et au chloroforme !!
- (f) excédent surnageant éliminé à la pipette

**ANNEXE 7 : Comparaison de l'activité de différents promoteurs** (adapté d'après Lutz *et al.*, 1988)

Promoteurs	Activité promoteurs				Facteur répression	Activité promoteurs en milieu LB (en milieu M63)		Facteur répression	Activité promoteurs		Facteur répression	Référence				
	Réprimé		Dérérimé			Réprimé	Induit		Réprimé	Induit						
	UGal	U Pbla	U Gal	U Pbla									RLU.cell <sup>-1</sup> x 10 <sup>-4</sup>	RLU.cell <sup>-1</sup> x 10 <sup>-4</sup>		
P <sub>lac</sub>	30	0.006	27500	5,5	920	5,73 (20,8)	270 (1029)	47 (49)				(Lanzer & Bujard, 1988)				
P <sub>lacL8</sub>	15	0.003	2000	0,4	150							(Lanzer & Bujard, 1988)				
P <sub>lacUV5</sub>	150	0.03	17000	3,4	110							(Lanzer & Bujard, 1988)				
P <sub>lacL8UV5</sub>	2740	0.55	46500	9,3	17							(Lanzer & Bujard, 1988)				
P <sub>tac</sub>	1510	0.36	88000	17,6	50							(Guzman <i>et al.</i> , 1995)				
P <sub>con/03</sub>	3740	0.70	59000	11,8	17							(Lanzer & Bujard, 1988)				
P <sub>A1/03</sub>	4180	0.83	111000	22,2	27							(Lanzer & Bujard, 1988)				
P <sub>A1con/03</sub>	8890	1.80	59500	11,9	7							(Lanzer & Bujard, 1988)				
P <sub>N25/03</sub>	8160	1.5	38500	7,7	5							(Lanzer & Bujard, 1988)				
P <sub>A1/04</sub>	110	0.022	190500	38,1	1730							(Lanzer & Bujard, 1988)				
P <sub>A1con/04</sub>	220	0.044	84000	16,8	380							(Lanzer & Bujard, 1988)				
P <sub>N25/04</sub>	250	0.050	88000	17,6	350							(Lanzer & Bujard, 1988)				
P <sub>A1/04s</sub>	58	0.011	125500	25,1	2200							(Lanzer & Bujard, 1988)				
P <sub>A1/05</sub>	13700	2.7	161500	32,3	12							(Lanzer & Bujard, 1988)				
P <sub>A1con/05</sub>	8500	1.7	38500	7,7	5							(Lanzer & Bujard, 1988)				
P <sub>con/05</sub>	nd	2.5	74500	14,9	6							(Lanzer & Bujard, 1988)				
P <sub>N25/05</sub>	nd	3.5	74500	14,9	4							(Lanzer & Bujard, 1988)				
P <sub>A1/03/04</sub> ou P <sub>A1lacO-1</sub> (colE1)									30	10430	350	(Lutz & Bujard, 1997) (Newman & Fuqua, 1999) (Guzman <i>et al.</i> , 1995)				
P <sub>BAD</sub>	4		1641		410,25	0,52 (0,32)	71,2 (593)	136 (1855)								
		0,0014à			10 000 à											
P <sub>L</sub>	7 à 42	0,008	0,6	58	30 000								(Lanzer & Bujard, 1988)			
P <sub>T7</sub> (pBR ori)	500												(Sektas <i>et al.</i> , 2001)			
P <sub>T7</sub>	250-1500												(Sektas <i>et al.</i> , 2001)			
P <sub>lac/ara1</sub> (colE1)													7	12400	1770	(Lutz & Bujard, 1997)
P <sub>trc</sub> (trp-lac)			équivalent Ptac										(de Lorenzo <i>et al.</i> , 1993)			
			18000-													
P <sub>tac/lac</sub> (pBR ori)	140-220		21000		129							(Sektas <i>et al.</i> , 2001)				
P <sub>tac/lacUV5</sub> (pBR ori)	>Ptac		<Ptac									(Morales <i>et al.</i> , 1991)				

U Gal : unités expression beta galactosidase (Lanzer & Bujard, 1988), U AP : mesure taux p-nitrophényl phosphate hydrolysé (Guzman *et al.*, 1992), RLU: Relative Light Units mesure d'une activité luciférase (Lutz & Bujard, 1997), M63 : milieu nutritif minimum.

## ANNEXE 8 : Alignement de séquences sous clustalX pour la définition des sites de recombinaison pheR1 et pheR2

*Ralstonia metallidurans* CH34 ou *Cupriavidus metallidurans* (numéro d'accèsion sur le site Nucleotide de NCBI : [CP000352](https://www.ncbi.nlm.nih.gov/nuclot/CP000352)), anciennement nommé *Alcaligenes eutrophus* puis *Ralstonia eutropha* (Mergeay *et al.*, 2003), a été entièrement séquencée par Copeland *et al.* (2006) (non publié). Des membres de ce genre bactérien ont été isolés des biotopes industriels et pollués (Mergeay *et al.*, 2003) et la souche *Alcaligenes eutrophus* 335 est capable de métaboliser le phénol (Hughes & Bayly, 1983). *Ralstonia metallidurans* présente dans son génome un ensemble de gènes intervenant dans la métabolisation de composés aromatiques de type phénoliques dont le gène codant l'enzyme phénol hydroxylase (**Figure A8-1**).



**Figure A8-1 : Région nucléotidique 1 935 195 à 1 938 670 bp du génome de *R. metallidurans* CH34.**

Rmet\_1783: Région conservée de l'enzyme phénol hydroxylase,  
 Rmet\_1784: gène codant méthane/phénol/toluène hydroxylase,  
 Rmet\_1785: gène codant l'enzyme monooxygénase MmoB/DmpM,  
 Rmet\_1786: gène codant l'enzyme méthane/phénol/toluène hydroxylase,  
 Rmet-1787: gène codant la sous unité de l'enzyme phénol hydroxylase

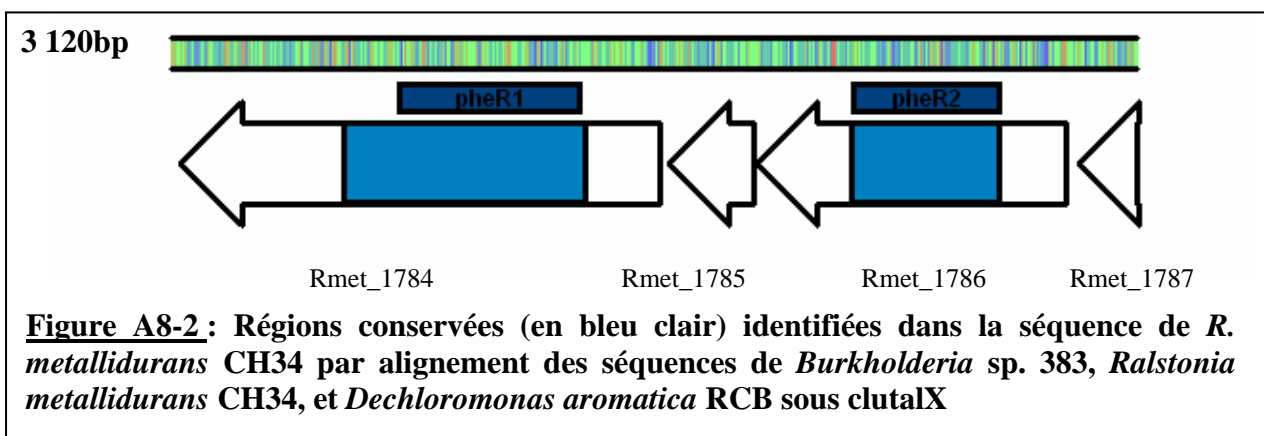
La séquence nucléotidique de la région codant la phénol hydroxylase (1 935 195 à 1 938 670 bp) du génome de *Ralstonia metallidurans* CH34 a été utilisée pour une analyse BLAST, afin d'identifier les espèces bactériennes présentant une similarité de séquences pour cette région (**Tableau A8-1**).

**Tableau A8-1 : Un exemple de résultat de l'analyse BLAST**

Accession	Description	Max score	E value
<a href="https://www.ncbi.nlm.nih.gov/nuclot/gi93352797 gb CP000352.1 ">gi93352797 gb CP000352.1 </a>	<i>Ralstonia metallidurans</i> CH34, compl	1568	0.0
<a href="https://www.ncbi.nlm.nih.gov/nuclot/gi57207881 dbj AB177762.1 ">gi57207881 dbj AB177762.1 </a>	<i>Wautersia numadzuensis</i> phtR, phtA	1017	0.0
<a href="https://www.ncbi.nlm.nih.gov/nuclot/gi72117119 gb CP000090.1 ">gi72117119 gb CP000090.1 </a>	<i>Ralstonia eutropha</i> JMP134 chromo	985	0.0
<a href="https://www.ncbi.nlm.nih.gov/nuclot/gi113528459 emb AM260480.1 ">gi113528459 emb AM260480.1 </a>	<i>Ralstonia eutropha</i> H16 chromosome	561	1,00E-156
<a href="https://www.ncbi.nlm.nih.gov/nuclot/gi4239910 dbj AB016860.1 ">gi4239910 dbj AB016860.1 </a>	<i>Pseudomonas putida</i> gene for pheno	547	2,00E-152
<a href="https://www.ncbi.nlm.nih.gov/nuclot/gi3445530 gb AF026065.1 AF02">gi3445530 gb AF026065.1 AF02</a>	<i>Ralstonia</i> sp. E2 positive p...	515	7,00E-143
<a href="https://www.ncbi.nlm.nih.gov/nuclot/gi5851787 dbj AB031996.1 ">gi5851787 dbj AB031996.1 </a>	<i>Ralstonia</i> sp. KN1 genes for regula...	478	2,00E-131
<a href="https://www.ncbi.nlm.nih.gov/nuclot/gi14275969 dbj AB051715.1 ">gi14275969 dbj AB051715.1 </a>	<i>Variovorax</i> sp. HAB-30 gene for ph...	476	6,00E-131
<a href="https://www.ncbi.nlm.nih.gov/nuclot/gi14275961 dbj AB051711.1 ">gi14275961 dbj AB051711.1 </a>	<i>Variovorax</i> sp. HAB-23 gene for ph...	460	4,00E-126

Accession	Description	Max score	E value
<a href="#">gi 14275965 dbj AB051713.1 </a>	Variovorax sp. HAB-27 gene for ph...	450	3,00E-123
<a href="#">gi 14275963 dbj AB051712.1 </a>	Variovorax sp. HAB-24 gene for ph...	450	3,00E-123
<a href="#">gi 14275967 dbj AB051714.1 </a>	Variovorax sp. HAB-29 gene for ph...	434	2,00E-118
<a href="#">gi 14275959 dbj AB051710.1 </a>	Variovorax sp. HAB-22 gene for ph...	424	2,00E-115
<a href="#">gi 17826934 dbj AB063252.1 </a>	Burkholderia kururiensis PhkR gen...	363	6,00E-97
<a href="#">gi 14275947 dbj AB051704.1 </a>	Ralstonia sp. HAB-01 gene for phe...	339	9,00E-90
<a href="#">gi 14276001 dbj AB051731.1 </a>	Uncultured microorganism PCRTD14	331	2,00E-87
<a href="#">gi 14276003 dbj AB051732.1 </a>	Uncultured microorganism PCRTD07	329	9,00E-87
<a href="#">gi 68250345 gb DQ026294.1 </a>	Pseudomonas sp. M1 phenol hydrox	323	5,00E-85
<a href="#">gi 4239908 dbj AB016859.1 </a>	Pseudomonas putida gene for phenol	321	2,00E-84
<a href="#">gi 14276005 dbj AB051733.1 </a>	Uncultured microorganism PCRTD08	315	1,00E-82
<a href="#">gi 14275955 dbj AB051708.1 </a>	Ralstonia sp. HAB-18 gene for phe...	315	1,00E-82
<a href="#">gi 14275957 dbj AB051709.1 </a>	Variovorax sp. HAB-21 gene for ph...	309	8,00E-81
<a href="#">gi 4239906 dbj AB016858.1 </a>	Pseudomonas putida gene for phenol	309	8,00E-81
<a href="#">gi 14275949 dbj AB051705.1 </a>	Ralstonia sp. HAB-02 gene for phe...	307	3,00E-80
<a href="#">gi 14275999 dbj AB051730.1 </a>	Burkholderia cepacia gene for phen...	299	8,00E-78
<a href="#">gi 4239912 dbj AB016861.1 </a>	Burkholderia cepacia gene for phen...	299	8,00E-78
<a href="#">gi 14276025 dbj AB051743.1 </a>	Uncultured microorganism PCRLD01	293	5,00E-76
<a href="#">gi 77968738 gb CP000152.1 </a>	Burkholderia sp. 383 chromosome 2,	291	2,00E-75
<a href="#">gi 14276019 dbj AB051740.1 </a>	Uncultured microorganism PCRLD18 .	291	2,00E-75
<a href="#">gi 11878080 gb AF282897.1 </a>	Burkholderia cepacia JS150 tb	289	7,00E-75
<a href="#">gi 14276023 dbj AB051742.1 </a>	Uncultured microorganism PCRLD08 .	285	1,00E-73
<a href="#">gi 72120635 gb CP000091.1 </a>	Ralstonia eutropha JMP134 chromos	281	2,00E-72
<a href="#">gi 14276021 dbj AB051741.1 </a>	Uncultured microorganism PCRLD06 ..	278	3,00E-71
<a href="#">gi 1008895 gb L40033.1 </a>	Pseudomonas sp. toluene/benze...	274	4,00E-70
<a href="#">gi 84180485 gb DQ342021.1 </a>	Pseudomonas sp. JF-10 phenol hydro.	270	7,00E-69
<a href="#">gi 46487982 gb AY531530.1 </a>	Wautersia taiwanensis phenol hydro...	262	2,00E-66
<a href="#">gi 71845263 gb CP000089.1 </a>	Dechloromonas aromatica RCB, compl	252	2,00E-63
<a href="#">gi 4239904 dbj AB016857.1 </a>	Pseudomonas putida gene for phen...	240	6,00E-60
<a href="#">gi 4239898 dbj AB016854.1 </a>	Pseudomonas putida gene for phen...	240	6,00E-60

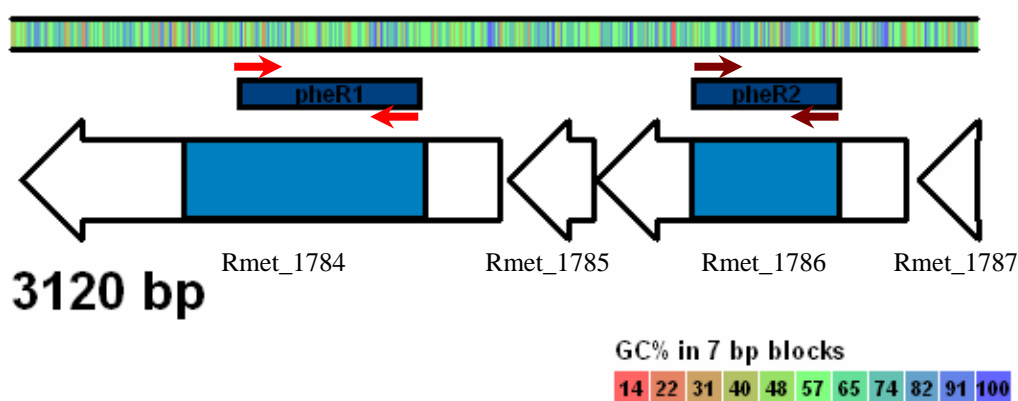
Les séquences de *Burkholderia* sp. 383, *Ralstonia metallidurans* CH34, et *Dechloromonas aromatica* RCB ont été choisies et alignées sous clustalX. Deux régions conservées ont été identifiées (**Figure A8-2**)



Des amorces pour l'amplification de pheR1 et pheR2 (**Figure A8-2**) ont été définies.

## ANNEXE 9 : Conditions d'amplification pour l'obtention de pheR1 et pheR2

Les couples d'amorces PheR1f / PheR1r et PheR2f / PheR2r définies avec oligo 6.1 à partir de la séquence de *Ralstonia metallidurans* CH34 (**Figure A9-1**) sont utilisés pour l'amplification des régions PheR1 et PheR2 qui seront utilisées comme site de recombinaison dans l'outil de capture de l'ADN de la microflore tellurique.



**Figure A9-1 :** Emplacement des couples d'amorces PheR1f/PheR1r (rouge) et PheR2f/PheR2r (rouge foncé) sur le génome de *R. metallidurans* CH34 ([CP000352](#))

Rmet\_1784: gène codant méthane/phénol/toluène hydroxylase,

Rmet\_1785: gène codant l'enzyme monooxygénase MmoB/DmpM,

Rmet\_1786: gène codant l'enzyme méthane/phénol/toluène hydroxylase,

Rmet\_1787: gène codant la sous unité de l'enzyme phénol hydroxylase

### Conditions d'amplification définies pour les couple d'amorces PheR1f/ PheR1r et PheR2f/ PheR2r

Le mélange réactionnel est préparé selon les indications données dans le chapitre III-1 avec 0,2 µL ADN métagénomique (extrait de biofilm d'un bioréacteur de traitements d'eaux usées) comme cible d'amplification par mélange réactionnel de 25 µL. Les amplifications sont effectuées par une dénaturation initiale à 95°C 3 min, suivi de 30 cycles constitués chacun des étapes de dénaturation à 95°C 1 min, d'hybridation à T<sub>m</sub> 1 min (voir sous chapitre III-2), puis d'élongation 72°C 1 min30. Une élongation finale de 72°C est effectuée pendant 10 min. La quantité de produit PCR obtenu avec le couple d'amorces PheR1f/PheR1r étant très faible, une amplification supplémentaire est réalisée sur le produit PCR purifié selon 15 cycles de dénaturation 95°C 1 min, hybridation 50°C 1 min et élongation 72°C 1 min. Les étapes de dénaturation initiale et élongation finale sont conservées.



## **ANNEXE 10 : Analyse par BLAST des produits d'amplification pheR1 et pheR2**

Les produits PCR pheR1 et pheR2 clonés dans pPR13 ont été amplifiés par le couple d'amorces Bad342Fw1/Bad342Rv1 et envoyés à séquencer. Les résultats du séquençage ont été analysés par BLAST.

**TableauA10-1 : Résultats Blast de la séquence pheR1 clonée dans le plasmide pPR13**

Accession	Description	Score (bit)	E value
gi 4239915 dbj BAA74747.1	phenol hydroxylase alpha subunit	109	6,00E-38
gi 14275998 dbj BAB58930.1	phenol hydroxylase large subunit ...	101	5,00E-35
gi 14276040 dbj BAB58951.1	phenol hydroxylase large subunit ...	103	2,00E-34
gi 57207886 dbj BAD86550.1	phenol hydroxylase component [Wauter	96.7	3,00E-34
gi 73541391 ref YP_295911.1	Methane/phenol/toluene hydroxyla...	96.7	3,00E-34
gi 94310722 ref YP_583932.1	methane/phenol/toluene hydroxyla...	95.5	5,00E-34
gi 4239907 dbj BAA74743.1	phenol hydroxylase alpha subunit [Pse	103	5,00E-34
gi 58201133 gb AAW66690.1	phenol hydroxylase large subunit [Pse	102	5,00E-34
gi 58201137 gb AAW66692.1	phenol hydroxylase large subunit [...	102	5,00E-34
gi 58201139 gb AAW66693.1	phenol hydroxylase large subunit [Pse	102	5,00E-34
gi 58201135 gb AAW66691.1	phenol hydroxylase large subunit [Pse	103	1,00E-33
gi 58201155 gb AAW66701.1	phenol hydroxylase large subunit [Pse	103	1,00E-33
gi 84180486 gb ABC54690.1	phenol hydroxylase alpha subunit [Pse	103	2,00E-33
gi 4239909 dbj BAA74744.1	phenol hydroxylase alpha subunit [Pse	103	3,00E-33
gi 14276012 dbj BAB58937.1	phenol hydroxylase large subunit ...	103	3,00E-33
gi 14276014 dbj BAB58938.1	phenol hydroxylase large subunit ...	103	3,00E-33
gi 14276038 dbj BAB58950.1	phenol hydroxylase large subunit ...	103	3,00E-33
gi 14275994 dbj BAB58928.1	phenol hydroxylase large subunit [Ps	103	3,00E-33
gi 58201179 gb AAW66713.1	phenol hydroxylase large subunit [Ps	102	3,00E-33
gi 58201169 gb AAW66708.1	phenol hydroxylase large subunit [...	102	3,00E-33
gi 58201165 gb AAW66706.1	phenol hydroxylase large subunit [...	102	3,00E-33
gi 14275964 dbj BAB58913.1	phenol hydroxylase large subunit [Va	100	3,00E-33
gi 14275968 dbj BAB58915.1	phenol hydroxylase large subunit [Va	100	3,00E-33
gi 14275966 dbj BAB58914.1	phenol hydroxylase large subunit [Va	100	3,00E-33
gi 14276048 dbj BAB58955.1	phenol hydroxylase large subunit ...	103	3,00E-33
gi 68250348 gb AAAY88747.1	phenol hydroxylase subunit [Pseudomon	103	4,00E-33
gi 4239905 dbj BAA74742.1	phenol hydroxylase alpha subunit [Pse	103	4,00E-33
gi 4239899 dbj BAA74739.1	phenol hydroxylase alpha subunit [Pse	103	4,00E-33
gi 14276002 dbj BAB58932.1	phenol hydroxylase large subunit ...	95.9	4,00E-33
gi 14276010 dbj BAB58936.1	phenol hydroxylase large subunit ...	103	4,00E-33
gi 14276008 dbj BAB58935.1	phenol hydroxylase large subunit ...	103	4,00E-33
gi 4239903 dbj BAA74741.1	phenol hydroxylase alpha subunit [Pse	103	4,00E-33
gi 14276022 dbj BAB58942.1	phenol hydroxylase large subunit ...	95.9	4,00E-33
gi 14276034 dbj BAB58948.1	phenol hydroxylase large subunit ...	103	4,00E-33

Accession	Description	Score bit	E value
gi 14275992 dbj BAB58927.1	phenol hydroxylase large subunit [Ps	103	4,00E-33
gi 58201161 gb AAW66704.1	phenol hydroxylase large subunit [	103	4,00E-33
gi 58201159 gb AAW66703.1	phenol hydroxylase large subunit [Pse	103	4,00E-33
gi 58201157 gb AAW66702.1	phenol hydroxylase large subunit [Pse	103	4,00E-33
gi 14275974 dbj BAB58918.1	phenol hydroxylase large subunit [Pse	103	4,00E-33
gi 14276046 dbj BAB58954.1	phenol hydroxylase large subunit ...	103	4,00E-33
gi 14275990 dbj BAB58926.1	phenol hydroxylase large subunit [Pse	103	4,00E-33
gi 14276036 dbj BAB58949.1	phenol hydroxylase large subunit ...	102	7,00E-33
gi 71909406 ref YP_286993.1	Methane/phenol/toluene hydroxyla...	94,7	1,00E-32
gi 14275962 dbj BAB58912.1	phenol hydroxylase large subunit ...	98,6	1,00E-32
gi 58201185 gb AAW66716.1	phenol hydroxylase large subunit	102	1,00E-32

**TableauA10-1 : Résultats Blast de la séquence pheR2 clonée dans le plasmide pPR13**

Accession	Description	Score bit	E value
gi 32967105 gb AAP92390.1	LapL [Pseudomonas sp. KL28]	239	4,00E-62
gi 67154618 ref ZP_00416363.1	Methane/phenol/toluene hydroxy...	235	5,00E-61
gi 29647418 dbj BAC75402.1	butylphenol hydroxylase component	231	1,00E-59
gi 110592636 ref ZP_01380992.1	Methane/phenol/toluene hydrox...	230	2,00E-59
gi 2338425 gb AAB67106.1	CrpB [Ralstonia pickettii]	228	8,00E-59
gi 11878082 gb AAG40789.1	Tbc1B monooxygenase [Burkholderia c	226	4,00E-58
gi 73539503 ref YP_299870.1	Methane/phenol/toluene hydroxyla...	223	4,00E-58
gi 111613218 ref ZP_01400300.1	methane/phenol/toluene hydrox...	222	5,00E-57
gi 1008897 gb AAA88457.1	alpha subunit-terminal oxygenase comp	220	2,00E-56
gi 17826937 dbj BAB79280.1	phenol hydroxylase subunit PhkB [Bur	219	5,00E-56
gi 75993176 gb ABA33697.1	phenol hydroxylase component [Burkh	209	3,00E-53
gi 94310272 ref YP_583482.1	methane/phenol/toluene hydroxyla...	209	3,00E-53
gi 67549269 ref ZP_00427138.1	Methane/phenol/toluene hydroxy...	209	4,00E-53
gi 110672104 gb ABG82167.1	BtxB [Ralstonia sp. PHS1]	207	2,00E-52
gi 71909408 ref YP_286995.1	Methane/phenol/toluene hydroxyla...	203	2,00E-51
gi 78063817 ref YP_373725.1	Methane/phenol/toluene hydroxyla...	202	4,00E-51
gi 73541389 ref YP_295909.1	Methane/phenol/toluene hydroxyla...	197	1,00E-49
gi 94310724 ref YP_583934.1	Methane/phenol/toluene hydroxyla...	194	2,00E-48
gi 3845546 dbj BAA34170.1	phenol hydroxylase component [Comam	191	1,00E-47
gi 47572110 ref ZP_00242156.1	hypothetical protein Rgel02003564	189	5,00E-47
gi 6505657 dbj BAA87869.1	phenol hydroxylase subunit [Comamon	186	3,00E-46
gi 57207884 dbj BAD86548.1	phenol hydroxylase component [Wauter	186	4,00E-46
gi 68250350 gb AAY88749.1	phenol hydroxylase subunit [Pseudomon	183	2,00E-45
gi 116694491 ref YP_728702.1	phenol hydroxylase P1 protein [...	182	6,00E-45
gi 3445533 gb AAC32453.1	phenol hydroxylase component; PoxB	181	8,00E-45
gi 5851790 dbj BAA84119.1	phenol hydroxylase component [Ralston	181	8,00E-45
gi 51242281 gb AAT99083.1	DmpL [Pseudomonas putida]	178	9,00E-44
gi 527548 emb CAA56741.1	subunit of phenolhydroxylase [Pseudom	178	9,00E-44

Accession	Description	Score bit	E value
gi 483479 emb CAA55661.1	phenol hydroxylase [Pseudomonas puti	178	9,00E-44
gi 1093587 prf 2104263B	catechol dioxygenase	178	9,00E-44
gi 118692 sp P19730 DMPL_PSEUF	Phenol hydroxylase P1 protein ...	178	9,00E-44
gi 67154181 ref ZP_00415926.1	Methane/phenol/toluene hydroxy...	178	9,00E-44
gi 47572192 ref ZP_00242238.1	hypothetical protein Rgel02003647	170	2,00E-41
gi 28848923 gb AAO47356.1	phenol hydroxylase component pH <sub>L</sub> [Pse	166	4,00E-40
gi 39931049 sp Q7WTJ6 DMPL_ACICA	Phenol hydroxylase P1 protei...	166	5,00E-40
gi 535281 emb CAA85381.1	phenolhydroxylase component [Acinet	158	1,00E-37
gi 2605609 dbj BAA23331.1	oxygenase component [Acinetobacte	157	1,00E-37
gi 90416275 ref ZP_01224207.1	Methane/phenol/toluene hydroxy...	151	1,00E-35
gi 13940416 emb CAC38006.1	hypothetical protein [Pseudomonas sp	96.7	3,00E-19
gi 73539497 ref YP_299864.1	Methane/phenol/toluene hydroxyla...	68.2	1,00E-10
gi 505050 gb AAB09622.1	toluene-3-monooxygenase oxygenase...	60.1	4,00E-08
gi 11878092 gb AAG40798.1	Tbc2E monooxygenase [Burkholderia	59.3	6,00E-08
gi 51247499 pdb 1T0Q B	Chain B, Structure Of The TolueneO-...	56.2	5,00E-07
gi 45479225 gb AAS66664.1	beta hydroxylase [Pseudomonas mendo	54.3	2,00E-06
gi 1174723 sp Q00460 TMOE_PSEME	Toluene-4-monooxygenase syste...	54.3	2,00E-06
gi 47575125 ref ZP_00245160.1	hypothetical protein Rgel02000372	52.8	6,00E-06
gi 94310254 ref YP_583464.1	methane/phenol/toluene hydroxyla...	52.0	1,00E-05
gi 78693151 ref ZP_00857665.1	hypothetical protein BradDRAFT...	51.6	1,00E-05
gi 47573392 ref ZP_00243431.1	hypothetical protein Rgel02002524	50.8	1,00E-05
gi 86738933 ref YP_479333.1	methane/phenol/toluene hydroxyla...	50.4	3,00E-05
gi 27378790 ref NP_770319.1	putative monooxygenase component...	49.7	5,00E-05
gi 10443295 emb CAC10509.1	beta-subunit of multicomponent te...	49.3	6,00E-05
gi 5911742 emb CAB55829.1	putative isoprene monooxygenase be...	48.1	1,00E-04

---

## **REFERENCES :**

**de Lorenzo, V., Eltis, L., Kessler, B. & Timmis, K. N. (1993).** Analysis of *Pseudomonas* gene products using lacIq/P<sub>trp</sub>-lac plasmids and transposons that confer conditional phenotypes. *Gene* **123**, 17-24.

**Guzman, L. M., Barondess, J. J. & Beckwith, J. (1992).** FtsL, an essential cytoplasmic membrane protein involved in cell division in *Escherichia coli*. *J Bacteriol* **174**, 7716-7728.

**Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. (1995).** Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177**, 4121-4130.

**Hughes, E. J. & Bayly, R. C. (1983).** Control of catechol meta-cleavage pathway in *Alcaligenes eutrophus*. *J Bacteriol* **154**, 1363-1370.

**Lanzer, M. & Bujard, H. (1988).** Promoters largely determine the efficiency of repressor action. *Proc Natl Acad Sci U S A* **85**, 8973-8977.

**Lutz, R. & Bujard, H. (1997).** Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* **25**, 1203-1210.

**Mergeay, M., Monchy, S., Vallaey, T. & other authors (2003).** *Ralstonia metallidurans*, a bacterium specifically adapted to toxic metals: towards a catalogue of metal-responsive genes. *FEMS Microbiol Rev* **27**, 385-410.

**Morales, V. M., Backman, A. & Bagdasarian, M. (1991).** A series of wide-host-range low-copy-number vectors that allow direct screening for recombinants. *Gene* **97**, 39-47.

**Newman, J. R. & Fuqua, C. (1999).** Broad-host-range expression vectors that carry the L-arabinose-inducible *Escherichia coli* araBAD promoter and the araC regulator. *Gene* **227**, 197-203.

**Sektas, M., Hasan, N. & Szybalski, W. (2001).** Expression plasmid with a very tight two-step control: Int/att-mediated gene inversion with respect to the stationary promoter. *Gene* **267**, 213-220.